DOCUMENT RESUME

ED 185 765	EC 123 920
AUTHOR	Ysseldyke, James E., Ed.; Mirkin, Phyllis K., Ed.
TITLE	Proceedings of the Minnesota Roundtable Conference on Assessment of Learning Disabled Children.
INSTITUTION	Minnesota Univ., Minneapolis. Inst. for Research on Learning Disabilities.
SPONS AGENCY	Bureau of Education for the Handicapped (DHEW/OE), Washington, D.C.
REPORT NO	IRLD-Mono-8
PUB DATE	Apr 79
CONTRACT	300-77-0491
NOTE	159p.; See also EC 123 901-925.
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	*Decision Making; Elementary Secondary Education; Evaluation Methods; Intervention; *Learning Disabilities; *Program Evaluation; *Student Evaluation; Student Placement

ABSTRACT

Proceedings from the Minnesota Roundtable Conference on Assessment of Learning Disabled children include two major presentations reviewing research and reactions to those presentations from conference participants. J. Ysseldyke's presentation "Psychoeducational Assessment and Decision Making" reviews basic considerations underlying his research on the assessment-intervention process; describes some of the issues involved (such as definitional debates, use of tests for purposes other than those for which they were intended, and questions regarding the tests' technical adequacy); and outlines six areas of research (including computer simulation studies and ecological research on placement team decision making). Reactions to the paper are given by B. Keogh, L. Goodman, and R. Woodcock. Discussion highlights on the topic follow. In the second major presentation, "Behavioral Research Methodology as a Basis for the Formative Evaluation of Learning Disability Services," S. Deno outlines background factors and assumptions of a 3 year study on the effectiveness of teacher implemented systematic formative evaluation. T. Lovitt, P. Newcomer, and J. Jenkins provide reactions, with highlights of the discussion included. A summary of discussion comments concluding the conference is also provided. (CL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED185765

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

James Ysseldyke

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IRLD

Director: James E. Ysseldyke

Associate Director: Phyllis K. Mirkin

The Institute for Research on Learning Disabilities is supported by a contract (300-77-0491) with the Bureau of Education for the Handicapped, Department of Health, Education, and Welfare, U.S. Office of Education, through Title VI-G of Public Law 91-230. Institute investigators are conducting research on the assessment/decision-making/intervention process as it relates to learning disabled children. Research activities are organized into eight major areas:

I. Adequacy of Norm-Referenced Data for Prediction of Success

II. Computer Simulation Research on the Assessment/Decision-making/Intervention Process

III. Comparative Research on Children Labeled LD and Children Failing Academically but not Labeled LD

IV. Surveys on In-the-Field Assessment, Decision Making, and Intervention

V. Ethological Research on Placement Team Decision Making

VI. Bias Following Assessment

VII. Reliability and Validity of Formative Evaluation Procedures

VIII. Data-Utilization Systems in Instructional Programming

Additional information on these research areas may be obtained by writing to the Editor at the Institute.

The materials presented herein were prepared under government sponsorship. Contractors are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official position of the Bureau of Education for the Handicapped.

Monograph No. 8

PROCEEDINGS OF THE MINNESOTA ROUNDTABLE CONFERENCE ON ASSESSMENT OF LEARNING DISABLED CHILDREN

James E. Ysseldyke and Phyllis K. Mirkin
Editors

Institute for Research on Learning Disabilities
University of Minnesota

April, 1979

James Ysseldyke

Stanley Deno

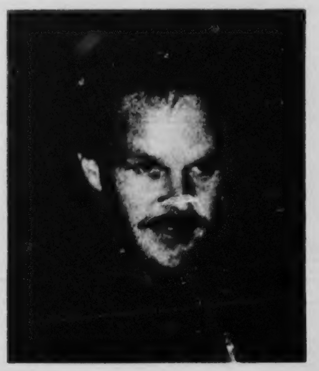
Barbara Keogh

John Guthrie

Phyllis Newcomer

Richard Woodcock

John Salvia


Joseph Jenkins

Thomas Lovitt

Libby Goodman (not pictured)

Table of Contents

Introduction
James E. Ysseldyke 1

Opening Remarks
Maynard C. Reynolds 9

Psychoeducational Assessment and Decision Making
James E. Ysseldyke 15

Psychoeducational Assessment and Decision Making: A Reaction
Barbara Keogh 37

Psychoeducational Assessment and Decision Making: A Reaction
Libby Goodman 43

Psychoeducational Assessment and Decision Making: A Reaction
Richard Woodcock 48

Follow-up Comments to Reactants
James E. Ysseldyke 52

Highlights of the Roundtable Discussion Following
Ysseldyke Presentation 54

Highlights of the Open Discussion Following
Ysseldyke Presentation 61

Behavioral Research Methodology as a Basis for the Formative
Evaluation of Learning Disability Services
Stanley L. Deno 65

Behavioral Research Methodology: A Reaction
Thomas Lovitt 91

Behavioral Research Methodology: A Reaction
Phyllis Newcomer 96

Behavioral Research Methodology: A Reaction
Joseph Jenkins 100

Follow-up Comments to Reactants
Stanley L. Deno 104

Highlights of the Roundtable Discussion Following
Deno Presentation 106

FEB 1 3 1980

Highlights of the Open Discussion Following Deno Presentation 110

Summary and Commentary John T. Guthrie 115

Summary and Commentary John Salvia 129

Highlights of the Roundtable Discussion Concluding the Conference 143

Highlights of the Open Discussion Concluding the Conference 148

Introduction

James E. Ysseldyke
University of Minnesota

The Institute for Research on Learning Disabilities at the University of Minnesota is one of five Institutes funded by the Division of Innovation and Development, Bureau of Education for the Handicapped. The Institute is focusing its research activities on critical and complex theoretical and empirical issues in the assessment-intervention and decision-making processes. Throughout the year preceding the Conference, Institute staff reviewed both the knowledge base and the state of the art in assessment and decision making, and outlined a plan to guide research activities during the next two years.

The Roundtable Conference afforded an opportunity for individuals with varied backgrounds in research, instruction, service, and test construction to react to what had already been done as well as to what was planned for the future; to share current perspectives on the state of the art and to propose future directions in research on the assessment of learning disabled children. The aim of the Conference was to create an atmosphere in which participants would be free to armchair--to speculate beyond current data and offer intuitive hunches about how best to address the critical needs in this area--to address issues in the context both of what has been and what might be.

Spring Hill, the site of the Conference, was an appropriate environment for exchange, speculation, and evaluation of ideas and views. The participants sitting at the Roundtable were persons who shared the Institute's serious investment in the improvement of assessment, decision making, and intervention for learning disabled children, who believed that through a cooperative, collaborative effort, giant strides could be made in that direction.

Unlike many other conferences, an audience was invited to eavesdrop and observe Roundtable exchanges, but more importantly to share ideas and reactions with the presenters during the two days of the Conference. Participants were urged to extend the Roundtable spirit beyond the conference room, to continue to communicate thoughts and concerns at coffee breaks, lunch, and dinner.

Two sessions on the first day of the Conference each began with one-hour presentations that reviewed the focal issues of the research and outlined, in general, the proposed activities of the research teams. The presentations derived directly from the previous year's efforts to review the knowledge base for current practices and to outline research believed necessary for addressing specific complex issues. Following each presentation, three individuals reacted to both the reviews and the research plans. Reactants were asked to address three issues: (1) the extent to which the principal investigators and their collaborators had adequately reviewed the current knowledge base from each of their particular perspectives, (2) the extent to which the rationale for doing research in a particular area was adequately developed, and (3) the potential payoff for the field, and learning disabled children in particular, of the planned research.

Following these presentations, individuals at the Roundtable were provided an opportunity to question the speakers for clarification or elaboration, to comment on specific issues of concern to them, and to discuss among each other those issues and concerns. They were able to, as it were, speculate, debate, and evaluate. After the Roundtable discussion, members of the participant-observer audience were given an opportunity to raise issues of concern to them and to ask questions of specific speakers.

The second day of the Conference began with comments by John Guthrie and John Salvia, who were asked to make some summarizing, integrative comments regarding the first day's discussions, as well as to offer some thoughts regarding overall direction in this area of research. Their remarks were followed by discussion among those at the Roundtable and, later, by comments and questions from the audience of participant-observers.

The Roundtable Conference and the preparation of this monograph were sponsored by funds made available from the Division of Innovation and Development, Bureau of Education for the Handicapped. The following are the individuals who participated in the Roundtable Conference:

Roundtable Participants

Robert Bruininks (Moderator)
Professor and Chairman, Department of Psychoeducational Studies
University of Minnesota, Minneapolis, Minnesota

Stanley Deno
Associate Professor, Department of Psychoeducational Studies
Principal Investigator, Institute for Research on Learning Disabilities
University of Minnesota, Minneapolis, Minnesota

Libby Goodman
Director, Special Education for Administrative Services
Stevens Administration Center, 13th and Spring Garden Street
Philadelphia, Pennsylvania

John Guthrie
Director of Research, International Reading Association
800 Barksdale Road, Newark, Delaware

Joseph R. Jenkins
Professor of Education, Special Education
Director, Experimental Education Unit
Child Development and Mental Retardation Center
University of Washington, Seattle, Washington

Barbara Keogh
Professor of Special Education, Department of Education
University of California, Los Angeles, California

Thomas C. Lovitt
Professor of Education, Special Education
Principal Investigator, Learning and Language Disabilities Training Project
Experimental Education Unit, Child Development and Mental Retardation Center
University of Washington, Seattle, Washington

Phyllis Mirkin
Research Associate, Department of Psychoeducational Studies
Associate Director, Institute for Research on Learning Disabilities
University of Minnesota, Minneapolis, Minnesota

Frederick J. Morrison
Visiting Associate Professor, Institute of Child Development
University of Minnesota, Minneapolis, Minnesota

Phyllis Newcomer
Associate Professor of Education
Beaver College, Glenside, Pennsylvania

Maynard Reynolds, Moderator
Professor, Department of Psychoeducational Studies
Director, Leadership Training Institute/Special Education
University of Minnesota, Minneapolis, Minnesota

John Salvia
Associate Professor of Special Education
Pennsylvania State University, University Park, Pennsylvania

Martha Thurlow
Assistant Scientist, Institute for Research on Learning Disabilities
University of Minnesota, Minneapolis, Minnesota

Richard Weinberg (Moderator)
Professor, Department of Psychoeducational Studies
Program Coordinator, Psychology in the Schools Program
University of Minnesota, Minneapolis, Minnesota

Richard Woodcock
Director, Measurement/Learning/Consultants
Tolovana Park, Oregon

James Ysseldyke
Associate Professor, Department of Psychoeducational Studies
Director and Principal Investigator, Institute for Research on Learning Disabilities
University of Minnesota, Minneapolis, Minnesota

Participant-Observers

Robert Algozzine
Consultant, Institute for Research on Learning Disabilities
Assistant Professor, Department of Special Education
University of Florida, Gainesville, Florida

Donald Allen
Graduate Research Assistant, Institute for Research on Learning Disabilities
University of Minnesota, Minneapolis, Minnesota

Gordon Alley
Professor of Special Education and Lecturer in Pediatrics
University of Kansas, Lawrence, Kansas

Marie Blackburn
Consultant, Institute for Research on Learning Disabilities
Coordinator, SIMS Center at Armatage, Department of Special Education
Minneapolis Public Schools, Minneapolis, Minnesota

Virginia Brown
Associate Professor of Special Education
University of Minnesota, Duluth, Minnesota

Tanis H. Bryan
Director, Chicago Institute for Learning Disabilities
University of Illinois, Chicago, Illinois

N. Dale Bryant
Director, Research Institute for the Study of Learning Disabilities
Teachers College, Columbia University, New York, New York

Berttram Chiang
Graduate Research Assistant, Institute for Research on Learning Disabilities
University of Minnesota, Minneapolis, Minnesota

Sandra Christenson
School Psychologist, The Groves Learning Center
Minnetonka, Minnesota

Deborah Clemmensen
Graduate Research Assistant, Institute for Research on Learning Disabilities
University of Minnesota, Minneapolis, Minnesota

Nicki Cohen
Graduate Research Assistant, Institute for Research on Learning Disabilities
University of Minnesota, Minneapolis, Minnesota

Robert Colligan
Associate Professor of Psychology
Mayo Medical School, Rochester, Minnesota

Pat Cronin
Control Data Corporation
8100 34th Avenue South, Minneapolis, Minnesota

Jan Bloom Davidson
Graduate Research Assistant, Institute for Research on Learning Disabilities
University of Minnesota, Minneapolis, Minnesota

Amos Deinard
Associate Professor, Department of Pediatrics
University of Minnesota, Minneapolis, Minnesota

Donald Deshler
Director, Research Institute for the Study of Learning Disabilities
University of Kansas, Lawrence, Kansas

Allen Dittmann
    Division of Innovation and Development
    U. S. Office of Education/BEH, Washington, D.C.

Bryon Egeland
    Professor, Department of Psychoeducational Studies
    University of Minnesota, Minneapolis, Minnesota

Carolyn Elliott
    Consultant, Specific Learning Disabilities and Emotional Disturbance
    Minnesota State Department of Education, St. Paul, Minnesota

Susan Epps
    Graduate Research Assistant, Institute for Research on Learning Disabilities
    University of Minnesota, Minneapolis, Minnesota

Patricia Evans
    Graduate Research Assistant, Institute for Research on Learning Disabilities
    University of Minnesota, Minneapolis, Minnesota

Nancy Fox
    Graduate Research Assistant, Institute for Research on Learning Disabilities
    University of Minnesota, Minneapolis, Minnesota

Marge Goldberg
    President, Parent Advocacy Coalition for Educational Rights (PACER)
    Minneapolis, Minnesota

Charles M. Hagen
    Director of Special Education, St. Paul Public Schools
    St. Paul, Minnesota

Daniel P. Hallahan
    Director, Learning Disabilities Research Institute
    Department of Special Education
    University of Virginia, Charlottesville, Virginia

J. Gary Hayden
    East Metro Special Education Regional Consultant (SERC)
    Inver Grove Heights, Minnesota

Raleigh J. Huizinga
    Executive Director, The Groves Learning Center
    Minnetonka, Minnesota

Michael Humphreys
    Chicago Institute for Learning Disabilities
    University of Illinois, Chicago, Illinois

Kathryn Kuehnle
    Graduate Research Assistant, Institute for Research on Learning Disabilities
    University of Minnesota, Minneapolis, Minnesota

Gary Lambour
    Division of Innovation and Development
    U. S. Office of Education/BEH, Washington, D.C.

Dennis LaRoque
    Director of Special Services, Duluth Public Schools
    Duluth, Minnesota

R. Eric Larson
Director, Special Education
Mower Education Service Cooperative, Brownsdale Elementary School
Brownsdale, Minnesota

Robert Lichtenstein
Graduate Research Assistant, Psychology in the Schools Training Program
University of Minnesota, Minneapolis, Minnesota

Lisa Lowry
Graduate Research Assistant, Institute for Research on Learning Disabilities
University of Minnesota, Minneapolis, Minnesota

Karen Lundholm
Assistant Director, Leadership Training Institute/Special Education
Department of Psychoeducational Studies
University of Minnesota, Minneapolis, Minnesota

Douglas Marston
Graduate Research Assistant, Institute for Research on Learning Disabilities
University of Minnesota, Minneapolis, Minnesota

James Moran
Graduate Research Assistant, Institute for Research on Learning Disabilities
University of Minnesota, Minneapolis, Minnesota

Lyelle Palmer
Department of Education
Winona State University, Winona, Minnesota

Stephen Poland
Graduate Research Assistant, Institute for Research on Learning Disabilities
University of Minnesota, Minneapolis, Minnesota

Margaret Potter
Graduate Research Assistant, Institute for Research on Learning Disabilities
University of Minnesota, Minneapolis, Minnesota

Marcie Radius
Castro Valley Child Service Demonstration Center
Castro Valley, California

Steve Ragan
Director of Special Education
Bemidji State University, Bemidji, Minnesota

Ronald E. Reeve
Learning Disabilities Research Institute
Department of Special Education
University of Virginia, Charlottesville, Virginia

Richard Regan
Graduate Research Assistant, Institute for Research on Learning Disabilities
University of Minnesota, Minneapolis, Minnesota

Mary Jo Richardson
Executive Director, Minnesota Association for Children with Learning Disabilities
St. Paul, Minnesota

Steven Robinson
Graduate Research Assistant, Institute for Research on Learning Disabilities
University of Minnesota, Minneapolis, Minnesota

Quentin Robley
Assistant Professor, Department of Special Education
St. Cloud State University, St. Cloud, Minnesota

Terry Rose
Child Service Demonstration Center
Northern Illinois University, DeKalb, Illinois

S. Jay Samuels
Professor, Department of Psychoeducational Studies
University of Minnesota, Minneapolis, Minnesota

Paula Savage
Graduate Research Assistant, Institute for Research on Learning Disabilities
University of Minnesota, Minneapolis, Minnesota

Margaret Jo Shepherd
Director, Learning Disabilities Training Program
Teachers College, Columbia University, New York, New York

Mark Shinn
Graduate Research Assistant, Institute for Research on Learning Disabilities
University of Minnesota, Minneapolis, Minnesota

Linda Stevens
Model Program Support System, The NETWORK
Andover, Massachusetts

Gerald Tindal
Graduate Research Assistant, Institute for Research on Learning Disabilities
University of Minnesota, Minneapolis, Minnesota

James E. Turnure
Professor, Department of Psychoeducational Studies
University of Minnesota, Minneapolis, Minnesota

John Wasson
Director, Learning Disabilities Training Program
Moorhead State University, Moorhead, Minnesota

Opening Remarks

Maynard C. Reynolds
University of Minnesota

This is an extraordinary period in Special Education. We are at the mountain top in the history of Special Education, at least in level of activity. This is the year in which the schools are mandated literally to locate and evaluate every handicapped child. It is the year in which individualized plans, millions of them, have had to be written - one by one - for all of these children. Truly, we have a great deal on the educational plate in 1978 as concerns handicapped students!

Events are running rapidly; in the movement of children from one structure to another, in the increasing studies of individual children, in after-school training sessions for teachers, and much more. There is a great press to get on with things very rapidly; legal rights are present rights. The procedural demands of new laws and regulations are very great.

The responses to all of these changes in many school districts and in the universities as well have been quite superficial. Most teachers have not been well prepared to write their IEPs.[1] Colleges have not geared themselves up to be helpful in the retraining processes. Some school leaders hope that most of the new activity will prove to be just another fad and disappear soon; and there is some downright hostility to the whole set of changes. Predictably, there will be attempts to take apart some of the legislation and regulations which now direct programs for the handicapped.

---

[1] Individualized Educational Programs - required for all handicapped students under Public Law 94-142.

It is noteworthy that all of the busy activities of the year are going on in a much more public way than ever before, so that where there is incompetency there is no hiding of it. Educational plans are being written by special teachers, regular teachers, and parents at meetings convened by school principals. As Nicholas Hobbs has remarked, it may be much like rediscovering Niagara Falls, but the schools are rediscovering parents. We are asking teachers to write carefully developed individual plans for students; they are not always competent to do that but all of it is being done in an unprecedented public way.

The field of learning disabilities, which we are considering at this conference, is in the middle of it all. It has been the most rapidly developing aspect of special education in recent years. Recently, it has drawn more attention politically than all other fields of special education combined. When there is vagueness or uncertainty about who these "learning disabled" youngsters are, or how we define this category of exceptionality, the policy makers are very concerned. It is the first category in view when concerns arise about limiting or "capping" the funds for special education.

Learning disabilities also gets attention because it deals with the most fundamental areas of the school curriculum - what Stoddard has termed the "cultural imperatives," such as language and basic mathematics. We carry a larger burden, perhaps, than any other field of special education in charting new ways, yet there are weaknesses in our situation which create much difficulty.

Let me mention just a few of the special areas of concern which come to mind in the field of learning disabilities. First, it may be observed that almost all of the problems of psychometrics or measurement come to the front in this field - problems of expectations, discrepancies, profiles, reliability, norms, and much more. Most definitions of learning disabilities start out with a statement about educability, usually involving IQ tests. So, there are all of the difficulties and burdens of that concept of educability. And then we proceed to noting (with only half an eye, since we pay attention to discrepancies in only one direction) whether or not the achievement of youngsters is up to what might be expected or predicted. I was reassured when Robert Thorndike said in his book Under-and Overachievement that it is hard to decide when the psychologist has overpredicted and when the child has underachieved. There are grave difficulties in dealing with discrepancy scores as part of a basic definition.

The field of learning disabilities, more than any other, is at the center of a major transformation in measurement systems away from what Leona Tyler calls a vertical emphasis, in which we emphasize simple kinds of predications about how high one might expect a person to go, toward a more horizontal emphasis, in which the concern is not for screening and sorting, but rather, on trying to design programs for the maximum development of each individual.

Second, there are problems of interdisciplinary communication. I am thinking of a very prominent article in the field of learning disabilities which on one page defines learning disabilities in terms of a neurological base and on the very next page indicates how lacking we are in knowledge about interactions between neurology and education. Until we get to the point of being able to specify such interactions the knowledge base at this point is empty! We have not learned how to scan or communicate well across disciplinary and professional lines and this is not a minor problem. The amount of money that goes into teaming arrangements, multi-disciplinary case studies, and multiple referrals is very great, yet the outcome is often very thin soup.

In order to develop good interdisciplinary communication, there must be concentration on one dominant focus at a time. I suggest that in the field of learning disabilities, the dominant focus should be on the teacher's problem, that is, on teaching and learning. We have not communicated very well to neighboring disciplines and professions about the decision making or the instructional problems of the teacher. I think this research institute faces a large challenge of helping to sort out the logic of judgments and decisions made in the classroom and to communicate some of that to other disciplines so that they can learn to communicate more effectively with teachers.

Third, there are big problems, moral as well as technical, with respect to when we begin treatment. There is a tendency in the field of learning disabilities to wait around until we have a full blown casualty - a big discrepancy - before beginning specialized treatment. We simply must seek the resources and structures by which we can identify children who need special attention earlier, differentiate programs earlier, and increase the rates at which children learn successfully in the areas of the cultural imperatives. Specialists must engage potential problems and not just full blown casualties. We must reduce the rate at which children experience years of failure in the schools.

However, when we move to earlier programs, we face again the problems of definition, identification, and labeling; we run head on into funding systems which require the big discrepancies. We haven't yet learned to deal with groups of children at risk and to meet problems of accountability on a statistical basis rather than on the basis of labeled individual children. As we proceed in studies of children with learning problems we must bring some of the knowledge back to the mainstream and help improve programs there. Some of our learning disability specialists could well be deployed into mainstream settings in addition to their clinical settings. As the Chilean poet, Mistral, has said, "Many things can wait, the child cannot. Right now his hip bones are being formed, his blood is being made, his senses are being developed. To him we cannot say 'tomorrow,' his name is today."

Fourth, there is the general problem of student classification. When the Congress asked educators to come up with the definition of "learning disability" some time ago, they made public a very considerable embarrassment - that we don't have a very good definition of the concept of learning disability. The systems we use now to define learning disabilities on the practical scene or to define educable or trainable mentally retarded are not all that real. As Paul Meehl says, the categories we use to define human behavior do not "carve nature at its joints"; they are, to a very considerable extent, political constructions. We ought to work at the general problem of human classifications and seek for better delineations, such as they may be needed for instructional purposes, in the schools.

I launched these few remarks by saying that in this year we have put a great deal on the educational plate concerning handicapped students. We are challenged as never before. Truly, it is a time for concentrated reflection as we hope will be achieved at this conference.

Psychoeducational Assessment and Decision Making[1]

James E. Ysseldyke
University of Minnesota

This paper is divided into three major sections. I'll describe some of the basic considerations underlying our research and take a look at why we are focusing on the assessment-intervention process, on assessment and decision making, and explain how we see those as being interrelated. Second, I'll describe some of the issues that have essentially troubled us regarding the assessment of learning disabled children, and talk about those issues as underlying premises for the kinds of research that we plan to be engaged in. Third, I'll describe the overall rationale or structure that is guiding some of our research activities and then describe, very briefly, those activities themselves.

Assessment in the Context of Decision Making

We are defining assessment, not in the traditional sense of "testing," but simply as the process of collecting data for the purpose of making decisions about pupils. In educational settings, there are essentially six kinds of decisions that are made using assessment data: referral decisions; screening decisions; classification, identification, or placement decisions; instructional planning decisions; pupil evaluation decisions; and program evaluation decisions. Now it is important from the outset, I think, to distinguish between those kinds of decisions because it will lead directly to some of the issues we face.

---

[1] This presentation is based on another paper: Ysseldyke, J.E. & Thurlow, M.L. Psychoeducational Assessment and Decision Making: A Review. In J. E. Ysseldyke and F. Morrison (Eds.), Multiple perspectives on assessment of learning disabled children, in press.

When we talk about referral, we talk simply about the identification of children for whom a referring agent believes there is sufficient difficulty that some specialist ought to take a look at the child. Screening, on the other hand, consists of the administration of tests, usually group tests, to groups of children for the purpose of identifying those who differ sufficiently from "normal," whatever that may be, that further assessment is believed to be warranted. A third, and very different, reason for engaging in the assessment of children is for the purpose of making classification, placement, eligibility, or identification decisions. In this case, our questions are really threefold. One, identifying the extent to which the child is handicapped; second, specifying the nature of the handicap; and third, identifying the least restrictive environment for the child. A fourth, related purpose, but again a very different reason for engaging in assessment, is for the purpose of planning either instructional or other kinds of interventions with children. Here, the questions we're asking in our assessment are twofold: We're trying to decide what to teach, which is a content question, and we're trying to decide how to teach, which is a question of the kinds of strategies, methods, techniques, etc., that will be effective with the child. I should mention that there are subsets of intervention planning; in some cases we're concerned with implementation, how to get an instructional program going, and in other cases we're concerned with adjustment, what to do in the process of intervention to modify or change things in order to move the pupil along. A fifth and very different reason for assessing children is for the purpose of evaluating the extent to which they are making progress in their educational programs. Parents, teachers, and children themselves have a very great need to know the extent to which progress is indeed being made. And, finally, assessment data are collected for the purpose of program evaluation, when we are attempting to evaluate the effectiveness of a particular instructional program. This activity is best evidenced in the kinds of evaluations we've seen of Head Start Programs and specific curricular interventions. The whole process of assessment is a decision-making process and one in which our assessment strategies and techniques ought to be dictated by the kinds of decisions we're trying to make.

The Use of Assessment Data to Make Intervention Decisions

Typically, traditionally, and too often currently, assessment and intervention are viewed as mutually exclusive activities. Educational personnel speak of assessment on the one hand, and intervention on the other, without viewing the two as integral parts of one dynamic process.

The effectiveness of any specific treatment or intervention is the function of a complex interaction of at least five identifiable factors:

1. The characteristics of the child,
2. The characteristics of the teacher,
3. The nature of the treatment or intervention employed,
4. The setting in which intervention is implemented,
5. The kind of behavior change we are trying to bring about.

Child characteristics interact with teacher characteristics, which in turn interact with the nature of the intervention used, with setting factors, and with the kind of behavior change we are attempting to bring about to affect the effectiveness of intervention efforts. We typically oversimplify a very complex process.

In spite of the complexity of factors affecting intervention effectiveness, educational personnel are charged with the task of deciding which of several alternative interventions to employ with a specific youngster. Such decisions should be data-based decisions.

Schools routinely collect many different kinds of data about children: scores on norm-referenced tests, data from criterion-referenced measures, observational data, interview data, medical information, developmental history data, social history information, and information regarding adaptive behavior. These data are used to make decisions, and the decisions themselves are interventions.

Figure 1 illustrates the use of data to make different intervention decisions. The large box at the top of the figure illustrates several different kinds of information or data that are either available or may be collected for the purpose of decision making. Certain data are used for the purpose of making screening decisions, other sets of data are used to make placement decisions, while still other data are used to make intervention decisions. As is illustrated in the figure, the same data may be used in making more than one kind of decision; more importantly, though, the data used do not entirely overlap. Different kinds of data are used for the purpose of making different kinds of decisions. Assessment and intervention are not static, they are dynamic parts of the assessment-intervention process. We believe we can best impact intervention by conducting research on the ways in which assessment data are used to make intervention decisions.

Insert Figure 1 about here

Definitional Debate

There is, and has been, little consensus among state departments in the ways in which they define learning disabilities. This fact is illustrated quite adequately by data summarized by Mercer, Forgnione, and Wolking (1975) and presented in Table 1. Considerable variance in the use of the Federal definition, in the specification of intellectual level, in the inclusion of process deficits, and in the inclusion/exclusion of emotional problems is readily apparent.

Insert Table 1 about here

I don't want us to spend the next two days debating alternative definitions of learning disabilities. That would be counter-productive. Numerous investigators before us have tried to identify causes of learning disabilities, and supposedly identified causes have run the gamut from spinal injury, developmental imbalances, and neurological dysfunction to inappropriate nurturance and instruction. Furthermore, investigators have been unable systematically and consistently to differentiate learning disabled children from either "normals" or other kinds of handicapped children. Clearly, major problems are evidenced in deciding the population about whom we are talking.

The Use of Tests for Purposes Other than Those for Which they were Designed

I indicated earlier that assessment data are used for making many different kinds of psychoeducational decisions. A major problem is the failure on the part of diagnostic personnel to differentiate their assessment strategies, devices, and techniques in light of the kind of decision to be made. What I'm referring to here is called the "WISC/Wide Range/Bender for every child, no matter what decision we're trying to make." This is probably best illustrated by the use of profile analyses in efforts to plan instructional interventions for children. Some of the best tests that are available are intellectual devices. They were originally designed for the purpose of helping us make classification and placement decisions. They will still do an extremely effective job of helping us make classification and placement decisions. But, we witness today individuals engaging in elaborate profile analyses using devices like the Wechsler Intelligence Scale for Children in efforts to plan instructional programs. If you go and pick up a psychological report, it'll read, "Johnny's poorest performance was demonstrated on a task requiring him to answer specific factual questions, while his best performance was demonstrated on a task requiring him to repeat sequences of orally presented digits. Johnny would profit from a program in which the teacher would have him answer many specific factual questions." Yet, there is no empirical support for that practice.

Technical Adequacy

Technically inadequate norm-referenced tests are too often used to gather data for the purpose of making important decisions. Three factors are involved. The first is standardization. When we assess individuals, we assume that the individuals we assess are like those in the normative group; that the task has been standardized on children who are like the child we're assessing. We assume that the individual has had a comparable set of experiences and opportunities to learn. A good number of the devices that we use on a daily basis to make decisions about children simply don't provide us with a set of norms. Table 2 lists tests with inadequately constructed or described norms.

Insert Table 2 about here

The second issue is one of reliability. Most of us, somewhere in our careers, took a measurement course, and we learned that assessment devices should be reliable. The commonly accepted standard is that a reliability of

.90 ought to be demonstrated before devices are used to make important decisions about individuals. Many of the devices that we use to make decisions about pupils lack the necessary reliability for effective use. Table 3 lists reliabilities of commonly used tests and indicates that for the most part individual and group intelligence and achievement measures do have reliabilities that are somewhere in the .90's. Reliabilities are all reported in ranges because they are taken directly from the test manuals and reflect differences across age. Reliabilities of measures of specific processes or abilities in no instance exceed .90. Such devices should not be used to make important decisions. The third issue is validity. In measurement courses we all learned a little ditty that says "Reliability is a necessary but not sufficient condition for validity." Devices that have inadequate reliabilities cannot be considered valid. Table 4 is a list of tests with questionable validity.

Insert Tables 3 and 4 about here

We've got some major problems when we start looking at the technical adequacy of the devices we use to gather data for decision making. I used to believe the notion that it's better to use a non-reliable test than to use no test at all. Yet, such tests are indeed dangerous because they afford the illusion that they are providing reliable, valid information.

Using Deficit Scores

The next issue is one of the use of deficit scores to identify the learning disabled. The reliability of a deficit score is nearly always lower than the reliability of either of the scores that go into it. So when we start using non-reliable devices, and then put those devices into an equation and compute the degree of deficit that a child demonstrates, the reliability we are dealing with is significantly low. Individuals have been able to show that reliance on the use of deficit scores results in a significant amount of mis-identification simply as a function of change.

Bias in Assessment

This nation has spent an enormous amount of time and effort in attempts to identify *the* test that is fair for use with members of specific racial or cultural groups. When I was in Pennsylvania, individuals in the State Department were engaged in the task of coming up with a list of tests that you could use with children who were members of specific racial groups. We haven't learned from a history of similar efforts in psychology. Psychologists have long debated the concept of "fairness," and have developed many mathematical models to compute the extent to which tests are fair. The one thing that is very clear to us at this point in time is that there is very little agreement among those who have talked about models of test fairness, and very little common consensus on the definition of non-discriminatory assessment.

We could suddenly have the fair test, a test that was fair for all children, and we would still have considerable bias in the assessment and decision-making process. Research has demonstrated that if you take psychological reports and put the picture of an attractive third grade child on the top of one psychological report, and put a picture of an unattractive third grade child on the other report, teachers and intervention agents make different kinds of recommendations and different kinds of prognoses for those children. People in decision-making situations do indeed discriminate or bias on the basis of things like sex, physical attractiveness, parental power within the system, SES, and so on.

Placement Team Decision Making

A set of issues is directly relevant to the placement team decision-making process. The point in time where assessment comes together is when a group of individuals sits down for the purpose of presenting data and tries to arrive at recommendations regarding the placement of the individual and the kind of instructional intervention that is necessary for that pupil. When we look at research relevant to the placement team decision-making process, there hasn't been a lot. There has been research on clinical judgment. Research on clinical judgment has typically employed a human information processing design, and has largely been inconclusive. The body of research that does exist indicates that individuals typically make very unreliable decisions. We face the task now of trying to see whether individuals who make unreliable decisions will suddenly achieve reliable and indeed valid decisions as a group.

A second body of research is research that has used a questionnaire-interview methodology. Yoshida, Kaufmann, Fenton, and their colleagues at the Bureau of Education for the Handicapped, conducted a series of investigations in Connecticut where they asked individuals who had participated on placement teams the extent to which they really participated and the extent to which they were satisfied with the process. Interestingly enough, that set of investigations showed that teachers were less satisfied with the whole process when they attended decision-making meetings than when they had been left out of them. In the first place, teachers felt that they didn't get an opportunity to participate. But the important thing, the only point I want to make, is that the only data we have from those investigations are data that give us useful information on perceived participation (how much you think you participated, how you felt about it), and the data are not direct data on the process itself.

A third line of investigation relevant to the placement team decision-making process has been a series of computer simulation studies, some of the reading research of Schulman and Vinsonhaler and their associates at Michigan State, Janet Lerner's program that is used to train individuals to come into consensus with expert opinion in diagnostic situations, and some of the research of Algozzine. In such studies, by means of the computer, people have been provided with data and then asked to make decisions about individuals. There are some real questions about the extent to which you can indeed simulate the decision-making process by means of a computer. One of the interesting sidelights, I think, to the research that we plan to do, is that we plan to study the same individuals by means of simulation, observation, and questionnaire/interview procedures. We may be able to get a handle on how much simulation really approaches reality in terms of telling us anything about the decision-making process itself.

The fourth line of investigation consists of use of systematic, naturalistic observation of the placement team decision-making process. That methodology, I think, is terribly attractive, but it's been relied on very little. We found only one investigation where people had gone in and looked at the placement team decision-making process. Patton (1976) conducted a study in which he looked at what happened, and he indicated that the placement team spent about five minutes per child in making a decision, and that the decision the school psychologist came in with was the decision that the team ended up making.

The last relevant bit of research has to do with decision-making models. We did survey the decision-making models; we found difficulty applying them to the kinds of questions that we're asking. There are industrial decision-making models, the group process models, where we look at how groups arrive at decisions, and so on. There is quite an extensive body of research on how teachers make decisions in the instructional process. Yet, we're having trouble finding models that fit our research.

Rationale for Research on Assessment and Decision Making

The foregoing review indicates clearly that there are many major questions regarding assessment and decision making, questions that are obviously worthy of investigation. We have chosen to engage in several lines of inquiry as a basic beginning for our longer-term research efforts.

It is our belief that the questions in this area can best be addressed by conducting several interrelated investigations. An overriding perspective guides our research on the placement team decision-making process: We believe that we must conduct our research in the naturalistic environment, endeavoring to achieve what Bronfenbrenner (1976) calls ecological and phenomenological validity.

Research on decision making to date has been largely contrived research. I don't know how many times in the last year I've received bogus case studies in the mail where people say, "Here are some data on a child; we'd like to have you make a decision about the child." The approach gets used so much now that people are becoming very sensitive to it. Individuals have looked at the influence of parents' marital status, race, sex, socioeconomic status, and so on, on the decisions that are made, but it's a contrived kind of situation. The case study, and in-lab simulations by means of computers, have been used to try to study real-life phenomena. The generalizability in those cases is presumed but never demonstrated. As a result, it's probably safe to say that we really know very little, if anything, about what really happens in decision making. We have some beliefs, some general ideas, and some pet theories. We've no ecologically valid data to support those beliefs, ideas, and theories. Such data, we believe, will be obtained by studying the decision-making process naturalistically in the environment in which it occurs, and by endeavoring to document and understand the many complex factors affecting the process.

But we also can't simply go out and observe behaviors and count behaviors that occur during the decision-making process; we have to strive for a second principle, what Bronfenbrenner (1976) calls phenomenological validity. He defines that by saying an ecological experiment cannot be solely behavioristic; provision must be made for assessing each participant's definition of the situation, how he or she perceives the setting, and its various elements. In naturalistic settings, far more apparently determines the behavior of decision makers than mere reliance on objective data. Decisions are made as a function of the characteristics of the children, to be sure, but decisions are also made as a function of setting, the beliefs, attitudes, motives, maybe even the reinforcement history of the decision makers. Decision makers have indicated that they've very often labeled youngsters as learning disabled in order to provide them with the services of a teacher that they believe is highly competent. Similarly, they state that they refused to label youngsters as learning disabled when they believed the services of a particular program or teacher were not in the best interests of the child. Merely observing the behavior of persons in decision making will not enable us to discover all of those factors we want to discover, it will be necessary to personally interview the decision makers to uncover those ecologically relevant factors. Thus, the research will employ several methodologies, where we not only go out and look at what happens, but we also ask people what happens, what they typically do in the process, and so forth.

Proposed Research

There are six lines of research in which we are involved. The first area of research looks at the adequacy of norm-referenced data for predicting success in a highly systematic instructional program. The research comes largely out of the observation that if you take a highly systematic instructional program, and you put children who are identified as learning disabled into it, a lot of children improve, but there are always a group of children for whom success comes only very slowly. Some of our questions in this research area are: Can we, on the basis of data that are available, at entry, differentiate children who are successful from children who are not? How do you define success in a systematic instructional program--in terms of number of objectives accomplished, in terms of return to a regular classroom situation, or what?

A second area of research is a series of computer simulation investigations. What we're concerned with here is looking at the assessment process itself as a function of the kinds of information presented to the assessor. The computer affords us the opportunity to systematically vary input data and to systematically vary the sex of the child, the physical attractiveness of the child, and so on, and then look at the decision-making process--look at the tests that individuals decide to use, look at the extent to which they access technically adequate data on the test, and look at, indeed, the kinds of outcomes that are achieved as a function of that process. We will simulate decision-making decisions, provide people with data, and ask them to go through the decision-making process itself. We also intend to look at that as a function of the knowledge base regarding assessment.

A third research area is really a two-fold study. We are conducting comparative research on children who are failing academically and who are labeled learning disabled and on children who are failing academically and who are not labeled learning disabled. How are those children different? Psychometrically and demographically, what are the differences in those two groups of children? We are really asking to what extent do those children demonstrate differences in their performances on task, to what extent do those children demonstrate SES differences, racial differences, sex differences, differences in physical attractiveness, differences in the power that individual's parents have within the school system? How do we differentiate the groups?

The fourth line of research is a series of questionnaire investigations where we ask decision makers how they make decisions. One investigation is designed to look at the ways building principals decide how to put children in different classes in the first place. Do people use rules to make those decisions, and if so, what are the rules that people use? Secondly, we intend to characterize the assessment process in the model demonstration centers, looking at the kinds of assessment devices and procedures that are used to make decisions and the extent to which those assessment procedures are differentiated as a function of decision making. In the third line of investigation, we will ask special education personnel to characterize the decision-making process for us. What happens, all the way from referral to outcome?

The fifth line of research will allow us to check out some of the information we receive from the other lines of research. It involves ecological research on placement team decision making. We intend to observe placement teams in the process of making decisions, and to look at the kinds of decisions that are made. We will look at the amount of time spent in presenting data, interpreting data, and also will contrast that with what people say happens in the decision-making process. We want to look at the kinds of data that are introduced and try to get a better handle on the process.

The last research area involves the notion of bias following assessment. We will look at the influence of having an LD sibling on later teacher behavior toward and expectations regarding an LD child, and we will look at the extent to which those expectations may differ as a function of social and academic considerations.

References

Bronfenbrenner, U. The experimental ecology of education. *Teacher's College Record*, 1976, *78*(2), 157-204.

Mercer, C. D., Forgnione, C., & Wolking, W. D. Definitions of learning disabilities used in the United States. *Journal of Learning Disabilities*, 1976, *9*, 376-386.

Patton, C. V. Selecting special students: Who decides? *Teacher's College Record*, 1976, *78*, 101-124.

Table 1

Number of States and Respective Percentages of Components Included in State Definitions

Components	No. of States	Percent
Definition		
NACHC only	9	21.4%
NACHC with variations	15	35.7%
Different	16	38.1%
None	2	4.8%
Intelligence		
Average and above	11	26.2%
Above mental retardation	8	19.1%
Not stated	23	54.8%
Process		
Process disorder	36	85.7%
Language disorder	35	83.3%
Academic		
Reading	31	73.8%
Writing	31	73.8%
Spelling	31	73.8%
Arithmetic	31	73.8%
Exclusion - primary		
Visual impairment	26	61.9%
Auditory impairment	26	61.9%
Motor impairment	23	54.8%
Mental retardation	21	50.0%
Emotional disturbance	25	59.5%
Environmental disadvantaged	23	54.8%
Exclusion - primary & secondary		
Visual impairment	3	7.1%
Auditory impairment	3	7.1%
Motor impairment	2	4.8%
Mental retardation	11	26.2%
Emotional disturbance	1	2.4%
Environmental disadvantaged	1	2.4%
Neurological impairment		
Included	4	9.5%
Not included	0	.0%
Possible	26	61.9%
Not stated	12	28.6%
Affective		
Includes emotionally disturbed	4	9.5%
Includes socially maladjusted	6	14.3%
Miscellaneous		
Attention deficits	5	11.9%
Motor deficits	7	16.7%
Thinking deficits	30	71.4%
Discrepancy component	12	28.6%
Special Education required	14	33.3%
Intraindividual differences	4	9.5%
Prevalence	2	4.8%
Chronological Age	4	9.5%

From Mercer, C., Forgnione, C., & Wolking, W. D. Definitions of learning disabilities used in the United States. Journal of Learning Disabilities, 1976, 9, 376-386.

Table 2

Tests with Norms That Are Inadequately Constructed or Described

Arthur Adaptation of the Leiter International Performance Scale (13)[a]
Bender Visual Motor Gestalt Test (15)
California Achievement Test (9)
Culture Fair Intelligence Tests (14)
Cognitive Abilities (14)
Developmental Test of Visual-Motor Integration (15)[b]
Developmental Test of Visual Perception (15)
Diagnostic Reading Scales (10)[b]
Durrell Analysis of Reading Difficulty (10)[b]
Full-Range Picture Vocabulary Test (13)[b]
Gates-MacGinitie Reading Tests (9)[b]
Gates-McKillop Reading Diagnostic Tests (10)[b]
Gilmore Oral Reading Test (10)
Goodenough Harris Drawing Test (14)
Gray Oral Reading Test (10)
Henmon-Nelson Tests of Mental Ability (14)
Illinois Test of Psycholinguistic Abilities (17)
Memory for Designs Test (15)
Metropolitan Achievement Test (9)
Peabody Picture Vocabulary Test (13)
Primary Mental Abilities Test (14)
Purdue Perceptual-Motor Survey (15)
Quick Test (13)
Silent Reading Diagnostic Tests (10)
Slosson Intelligence Scale (13)
Stanford-Binet Intelligence Scale (13)
Wide Range Achievement Test (9)

---

[a] Numbers in parentheses refer to the chapter in which the test is described.

[b] These tests include norms in their manuals but include no data about the group on whom the test was standardized.

Table 3

Reliabilities of Frequently Used Tests

Measure	Reliability
California Achievement Test (Subtest Reliabilities)	.76 - .97[a]
Iowa Test of Basic Skills (1974 edition)	None
Metropolitan Achievement Test	.84 - .96[c]
Stanford Achievement Test (1973 edition)	.65 - .97[a]
Gates-MacGinitie Reading Test	.88 - .96[c]
Peabody Individual Achievement Test	.42 - .94[b]
Wide Range Achievement Test	[c]
Gray Oral Reading Test	.97 - .98[d]
Gilmore Oral Reading Test	.53 - .94[d]
Gates-McKillop Reading Diagnostic Test	None
Durrell Analysis of Reading Difficulty	None
Stanford Diagnostic Reading Test (1976 edition)	.75 - .94[c]
Silent Reading Diagnostic Test	.85 - .97[c]
Diagnostic Reading Scales	.87 - .96[a]
Woodcock Reading Mastery Tests	.79 - .99[c]
Key Math	.39 - .90[a]
Stanford Diagnostic Mathematics Test	.84 - .97[a]
Stanford-Binet Intelligence Scale	None
Wechsler Intelligence Scale for Children--Revised	
Verbal	.91 - .96[c]
Performance	.89 - .91[c]
Full Scale	.95 - .96[c]
Subtests	.62 - .92[c]
Wechsler Adult Intelligence Scale	
Verbal	.96[c]
Performance	.93 - .94[c]
Full Scale	.97[c]
Subtests	.60 - .96[c]
Wechsler Preschool and Primary	
Verbal	.93 - .95[c]
Performance	.91 - .95[c]
Full Scale	.96 - .97[c]
Subtests	.62 - .91[c]
McCarthy Scales of Children's Abilities	
Verbal	.86 - .92[c]
Perceptual-Performance	.75 - .90[c]
General Cognitive	.90 - .94[c]
Quantitative	.77 - .86[c]
Memory	.72 - .83[c]
Motor	.60 - .84[c]

Table 3 (continued)

Measure	Reliability
Full Range Picture Vocabulary Test	None
Quick Test	.60 - .96[d]
Peabody Picture Vocabulary Test	.67 - .84[d]
Nebraska Test of Learning Aptitude	.92 - .95[c]
Blind Learning Aptitude Test	.93[a]
Arthur Adaptation of the Leiter International Performance Scale	None
Pictorial Test of Intelligence	.87 - .93[a]
Columbia Mental Maturity Scale	.85 - .91[a]
Culture Fair Intelligence Scale	
Scale 1 Total	.80[b]
Scale 2 Total	.71 - .81[a]
Scale 3 Total	.51 - .68[a]
Cognitive Abilities Test (Total)	.91 - .95[a]
Goodenough-Harris Drawing Test	.60 - .70[b]
Henmon-Nelson Intelligence Test (Total)	.84 - .97[a]
Kuhlmann-Anderson Intelligence Tests (Total)	.93 - .95[a]
Otis-Lennon Mental Ability Test (Total)	.88 - .96[a]
Primary Mental Abilities Test (Total)	.86 - .95[b]
Short Form Test of Academic Aptitude (Total)	.90 - .96[a]
Bender Visual Motor Gestalt Test (1975 manual)	.50 - .90[b]
Developmental Test of Visual Perception (Subtests)	.29 - .70[b]
Developmental Test of Visual Perception (Total)	.69[b]
Memory for Designs Test	.72 - .90[b]
Purdue Perceptual-Motor Survey	None
Goldman-Fristoe Test of Articulation	Inte-rater Only
Wepman Auditory Discrimination Test	None
Northwestern Syntax Screening Test	None
Illinois Test of Psycholinguistic Abilities	
Subtests	.12 - .90[b]
Total	.66 - .91[b]

[a] Internal Consistency

[b] Test-Retest

[c] Split-half

[d] Alternate form

Source: Data are from test manuals and were compiled from data reported earlier by Ysseldyke and Salvia (1974) and Salvia and Ysseldyke (1978).

Table 4

Tests Having Questionable Validity[a]

- Bender Visual Motor Gestalt Test
- California Achievement Test[b]
- Developmental Test of Visual-Motor Integration[b]
- Developmental Test of Visual Perception
- Durrell Analysis of Reading Difficulty[b]
- Full-Range Picture Vocabulary Test[b]
- Gates-MacGinitie Reading Tests
- Gates-McKillop Reading Diagnostic Tests[b]
- Gilmore Oral Reading Test[b]
- Gray Oral Reading Test[b]
- Henmon-Nelson Tests of Mental Ability
- Illinois Test of Psycholinguistic Abilities
- Metropolitan Achievement Test[b]
- Purdue Perceptual-Motor Survey
- Stanford-Binet Intelligence Scale[b]
- Wide Range Achievement Test

---

[a] Adapted from Salvia and Ysseldyke (1978).

[b] No validity data are included in the manuals for these tests.

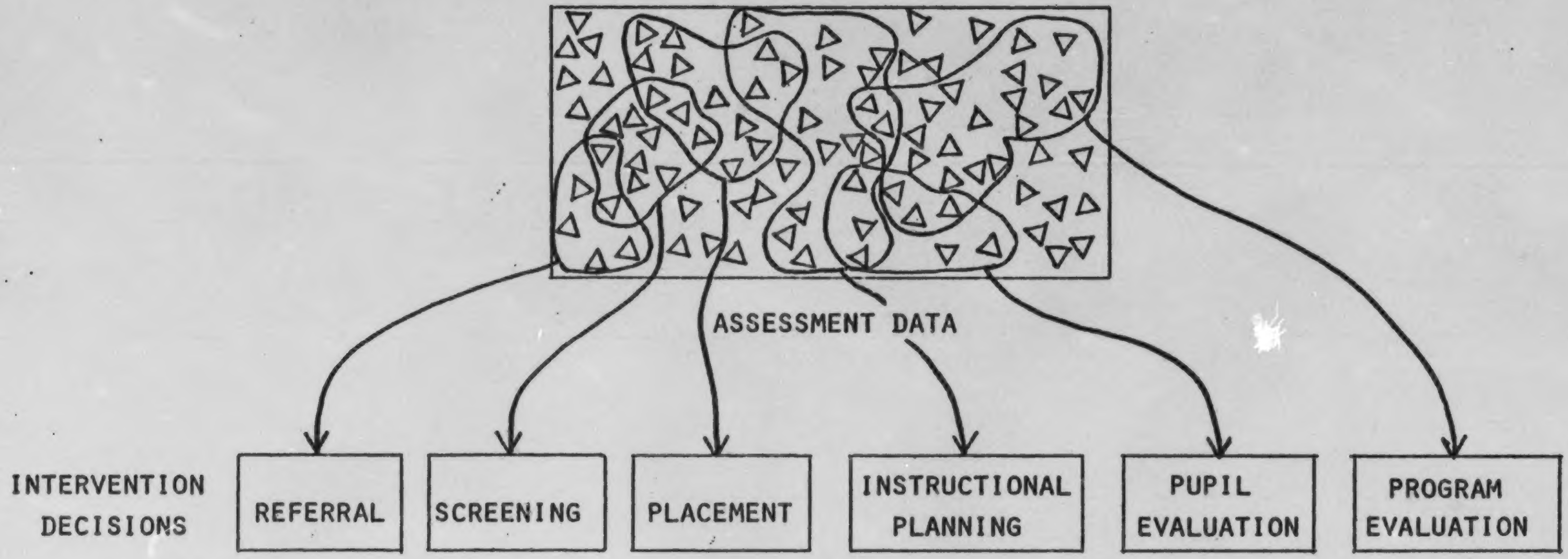

ASSESSMENT DATA
INTERVENTION DECISIONS
REFERRAL
SCREENING
PLACEMENT
INSTRUCTIONAL PLANNING
PUPIL EVALUATION
PROGRAM EVALUATION

Figure 1

Use of Data to Make Different Intervention Decisions

Psychoeducational Assessment and Decision Making: A Reaction

Barbara K. Keogh
University of California, Los Angeles

Research within this section of the Minnesota LD Institute is focused on two major lines of work. One involves inquiries directed at the characteristics of learning disabled children; another deals with the characteristics of decision makers and the decision-making process. I would like to emphasize my enthusiasm for these research directions, and to underscore the importance of systematic study of assessment procedures and of the people who make decisions about children. I was particularly impressed with the utilization of a variety of data sources. Few people study principals, teachers, school psychologists, *and* children. Most study only one of these groups. I commend you for broadening your base of information and for utilizing different kinds of information.

As I see it, the fundamental question being addressed in this series of studies is "What is the link between assessment data and educational decisions about LD children?" Answers to this question require knowledge about children and knowledge about decision makers. Thus, the research plans appear to be on target. What has not been made explicit, however, is the overall conceptualization of the linkage between the two sources of information. At the conclusion of this research program you will know a considerable amount about assessment, about the technical adequacy of instrumentation, and so forth. You will also know something about decision-making procedures. But will the data allow us to understand, and inferentially at least, to improve decision making *vis a vis* children with learning problems? It is to this point that I wish to address several brief remarks.

At least three well developed conceptual approaches to the study of decision processes are applicable to decision making within an educational setting. I refer specifically to the work of Richard Shavelson on teacher's decision making, to the work of Bernard Weiner on attribution, and to the social power research of Bertram Raven. Although the latter two theoretical approaches were developed within somewhat different contexts, all three seem useful for clarifying and understanding educational decision making.

Shavelson views teaching as a decision-making process. Based on work of other theorists, he has identified some heuristics which teachers use to synthesize the complex information and the variety of data available when making instructional decisions about children. Shavelson's heuristics are appropriate in the decision-making context in which your work is based. As example, Shavelson talks about a "representative" heuristic, an "availability" heuristic, and an "anchoring" heuristic. To illustrate the potential use of this approach in your work, the first heuristic implies that when given new information, we tend to accept and credit that which is consistent with our view of the topic. Specifically, if a youngster matches our view or is "representative" of our view of what mentally retarded children "look like," we are apt to accept information consistent with that diagnosis, even when the information is unreliable. Such an heuristic may help explain why some bits of unreliable information might be accepted and utilized in the decision-making process when others are discarded. Analysis of the utilization of information by decision committees in special education might be facilitated and clarified utilizing this heuristic. Importantly, understanding of this heuristic also ties the child's characteristics to the decision. The other two heuristics touch on somewhat different yet related aspects of selection and utilization of information in decision settings, but the overall point to be emphasized is that there are a number of ways information is utilized in decision making, and these processes need specification and study.

I would like to see you generate more information about the nature of the decision-making process in terms of the persons who make the decisions. You have mentioned that we use data to make decisions, and that the decisions are influenced by the nature of the data collected. I would argue further that the data you choose, and how you interpret them are less a function of the technical adequacy of the instrument than of the "set" of the person who is doing the selection and making the decision. This aspect of the decision-making process is deserving of attention and of systematic investigation.

In this regard, the work of Bernard Weiner on attribution also seems directly related to what you are doing. Weiner has proposed that in synthesizing information about individuals, we tend to make attributions about the causes of their condition. When a child has been referred as "having some sort of failure in school," we immediately begin to seek reasons for that failure. Weiner has proposed that in most achievement situations, we make attributions to stable or unstable causes and to internal or external ones. We may say that a child is not doing well because he or she is not very smart, thus making an attribution to ability -- an internal, stable cause. If one views a child's failure in school as being due to lack of ability, then the obvious thing to change is the nature of the program, the curriculum, or the placement, i.e., put the child in a special class, or somehow to change the educational environment to be consistent with the child's presumed ability level. On the other hand, if we make an attribution to motivation or to some less stable kind of characteristic, we are apt to attempt to change the child, i.e., to get the child to "try harder." The point is that teachers make decisions about what they are going to do in terms of their attributions about causes of a child's disturbance or problems. I suspect that decision-making teams make the same kinds of attributions in terms of stable-unstable, internal-external characteristics as proposed in Weiner's model. Thus it seems the attribution approach might provide some power in identifying the dynamics that go on within the decision-making team.

We have begun to explore this in some of our work at UCLA, and I can attest to the power of this model. In a study by Lavelle, we provided achievement performance information about their own youngsters to parents of mentally retarded, learning disabled, and normal children. When we asked the parents of normal children why their children were successful or unsuccessful on a particular task, the answers were as you might expect. When the child was successful, the parents said, "He does well on that kind of assignment." If the child was unsuccessful, the parent tended to write that off and said, "Well, he had a bad day," or, "He didn't understand that particular question." When we asked the parents of learning disabled children the same questions, they made motivational attributions in almost all cases. If a child did well, they said, "He's really a smart kid, if he'd only try," or, "He really worked hard on that day." If the child failed, the parents said, "That's one of his problems. We just can't get him to buckle down and really stay with something." The parents of the mentally retarded children, however, made what is really a very sad kind of attribution. When the child was unsuccessful, the parents said, "He is retarded and he never does very well on school tasks." When the child was successful, the parents were apt to write that off and say, "He must have been lucky," or, "It was a very easy kind of task." Understanding the nature of the attributions of decision makers is an important facet of your research. The attribution model seems particularly well suited to this topic.

The social power literature provides still another way of analyzing decision making. A social power model may provide some understanding of the interpersonal dynamics within decision-making teams. Raven and his colleagues, working within the social psychology field, have studied group processes in terms of social power. In their view, individuals influence other people on the basis of a number of possible dimensions -- coercive power, legitimate power, and referent power, to mention only three in the Raven model. "Social power" dimensions may provide insights into why some parents have more influence on school decisions than do others, or why certain members of the decision team are more effective than others. As example, do administrators accede to particular parents' requests because they view those parents as having some coercive power? Do school psychologists carry more power than other members in the diagnostic or placement teams because they are viewed as having "expert" status? Do decision makers make differential decisions about families and children whom they view as similar to themselves? Social power analyses may shed some light on the nature of the decision team interactions which result in decisions about children.

The three approaches or models I have mentioned are examples of ways of conceptualizing this interesting research area. While I consider these models potentially useful, there are clearly a number of others which provide powerful direction and organization to the research. The point to be emphasized is that an explicitly stated conceptualization is needed in order to direct the research activities and to integrate and interpret the findings of the various studies. More importantly, perhaps, the conceptualization will provide the bridge between the studies of assessment and decision making. It is this linkage which is the key to appropriate services for LD children. The Minnesota Institute is to be congratulated for taking on this complex problem.

Psychoeducational Assessment and Decision Making: A Reaction

Libby Goodman
Philadelphia Public Schools

I have reacted to the paper as a consumer, probably because of my ties to public school programming. As I was reading, I was looking for some indications of how the research at the Institute for Research on Learning Disabilities will help me deal with learning disabled youngsters. I believe my consumer perspective is consistent with the charge, given to the research institutes, to become involved in basic research that would ultimately lead to practical applications. The questions that Ysseldyke and his colleagues have posed, related to the linkage between assessment and decision making, have accomplished those ideals. The strong commitment to investigate the decision-making process, particularly within the context of the multiple disciplinary team, was most appropriate and very welcome to me personally. At the public school level, the team meeting has become a critically important point in the total process of referral, identification and placement. Individual professionals who once held sway over the identification and placement of learning disabled children, such as the psychologists, or the coordinator of the LD program, no longer have the power, if the team functions as it is supposed to, to arbitrarily or unilaterally make decisions that affect children's lives. The decisions which impact on our youngsters are the outcome of a multidisciplinary process and no longer are the sole prerogative of professionals alone. Parents and their advocates are also part of the team if they wish to be.

Unfortunately, some professionals resent the loss of that power. Others resent what they consider to be an interference. Perhaps some are annoyed by the added demands that are now made upon their time. I am hopeful that all school people will come to realize that, in this day of accountability and legalities, collective responsibility is indeed a very good thing.

Personally, I welcome it. It is important that parents also realize that their participation in the process carries with it a share of the responsibility. The placement decisions that are made, the program that is hammered out and documented in the IEP, represent the work, the wishes of all participants; we, the professionals and the parents, must share the glory or the blame. The functioning of the team has implications for the whole school system. Initially, it exposes us to public view and to public censure if that is necessary. We have to put our professionalism, our judgments, and our activities on the table. And the public can say, "Yes, we like that" or "No, we don't like that." If the multidisciplinary team functions well, the educational process proceeds as it should. If it doesn't, procedural safeguards come into play, children are inappropriately placed, educational programs are interrupted, confrontations emerge, legal advocates come into the picture, and so forth. The work of the Institute relating to the factors that contribute to decision making, the distinct roles of the members of the multidisciplinary teams, and so on, is going to be extremely valuable to us. The questions that you pose, certainly hint at the many practical applications; I urge you to proceed as quickly as possible. I'm very anxious to see the results and I have a feeling that they are going to be extremely helpful.

You mentioned the need to distinguish learning disabled from other types of handicapped. You opened with the plea that we not digress into a lengthy discussion of the definition of "learning disability." I do not want to get into the middle of that issue for it has no end. However, I feel compelled to say that I am concerned that the planned investigations, particularly those that relate to characteristics of learning disabled populations, may add little to our understanding of the nature of learning disabilities. Much of the proposed research rests heavily on the use of many of the tests that you berated for problems of reliability, validity, norming, and so forth. I wonder how much more we will know about the nature of learning disabilities from using such instruments. The issue of definition cannot be ignored. I am not proposing that the Institute undertake to put forth a new definition, but I would urge you to consider the possibility of expanding the scope of some of your investigations in order to delve into the identifying and distinguishing characteristics of learning disabled children.

Reference was made to the new regulations on learning disabilities. I suspect that many of us in this room realize that the new regulations really do not solve the underlying problems of definition and criteria. In one research area, an operational definition is used to identify the learning disabled students that would be the target population. Essentially, it stated that the children would be functioning two years below capability or expectancy. This is an underachievement model. This approach, which depends merely on a discrepancy between achievement and expectation, will not do. This operational definition will very quickly disintegrate within the context of an urban school, and under the pressures of an urban environment. By this standard, as many as 60 percent of the children of the entire school age population in certain subdistricts of Philadelphia would be learning disabled. The learning disabilities category, particularly in large cities, is abused and is subject to overinclusion. The abuse of the label is amplified in an urban setting. Learning disabilities often is confused with juvenile delinquency. Yes, there is learning disability and juvenile deliquency overlap, but not a one to one correspondence. Learning disabilities is confused with compensatory education. Learning disabilities is now seen as the alternative to EMR classification. Assessment must help us to distinguish the truly learning disabled from the larger mass of underachieving children. This issue cannot be ignored or set aside.

I agree with you that a prevalent approach in the past was to focus on identification of the cause, followed by remediation of the underlying cause, and finally, remediation of skill deficits - a) cause, b) remediation of cause, and c) remediation of academic problems. The results of such investigations have been generally disappointing. But I question whether that a-b-c sequence is unalterable. Does the identification of the cause necessarily dictate an effort at its remediation? This course of action would certainly be questionable if the causative factors are biologically based. Have we adequately researched underlying causes? I ask this question because I suspect that perhaps some of the differentiating factors or characteristics we are looking for lie in an investigation of the etiological factors and the biological bases of learning disabilities. Remediation of skill deficits, the last part of the a-b-c chain, has certainly been the most productive. It has given us immediate results and many classroom applications. The researchers and educators in that area have given us a technology that applies to all children with learning problems. And, it is a technology which addresses the issue of management of learning problems more so than the identification of learning disabilities. Note that I draw a distinction between the identification of the child and management of the learning problem.

One reference in your paper stated that learning disabilities is not a label which applies to some children, but rather a label which applies to all children and that the goal may well be to develop an exemplary instructional technology which is applicable across all exceptionalities. This is one alternative: essentially, the abandonment of the LD category as a distinct category of exceptionality. Although I personally believe that there are truly learning disabled youngsters, I would prefer this alternative to the present state of affairs and the confusion that still exists in our field. Unless we confront the problem of distinguishing learning disabled youngsters from the non-handicapped and the further problem of distinguishing the learning disabled from other categories of handicapped, I am afraid that this will be the only alternative and the ultimate outcome. Jim, the title of your paper was "Psychoeducational Assessment and Decision Making." I note that it excluded "for the learning disabled." Was this a premonition of the future?

Psychoeducational Assessment and Decision Making: A Reaction

Richard Woodcock
Measurement/Learning/Consultants

I'm going to assume the role of the gadfly. My comments don't concern major matters. I didn't find any. First of all, I want to talk about three of the terms that were used in your paper. I think that in the field of measurement and its application to learning disabilities, there are misconceptions about these terms which may mislead our thinking. One term I'm concerned about is used when talking about "norm-referenced tests" and "criterion-referenced tests." It is the word test that I'm concerned about. It seems to me that the term misdirects what it is that we're talking about. It is not the test that is criterion referenced or norm referenced, but rather certain forms of interpretation that we can apply to the results that we get from a test. The same test results for a certain test could be interpreted either way, or a test that was designed to serve one kind of assessment purpose might very well be used to serve some other sort of assessment purpose. It should be made clear that it is the interpretation we're talking about, not tests.

I am also concerned about the use of the word "standardization" in the paper as well as in the field. I do not think of the word "standardization" as being synonymous with "norming." Norming is part of the process of standardization. Norming may or may not be an essential ingredient in that process. Standardization, to me, is the process of insuring that the test will be administered, scored, and interpreted in the same way by different people. To insure consistent interpretation in norm referenced interpretation, tables will be needed. Norming is the way to obtain the numbers for those tables.

The third term that concerns me is "reliability." When we talk about reliability, we are talking about the precision of scores obtained from tests. I believe that people in the field assume that the two statistics "standard error of measurement" and "reliability coefficient," are just two different facets of the same thing. I don't believe they are at all. Basically, we have two different kinds of scores that come out of tests. We have scores that reflect the amount of an ability that a person has; these are called "raw scores" or "grade equivalent scores" or "age equivalent scores." With the advent of the Rasch model, we are now able to talk about some other kinds of scores that seem to have superior characteristics psychometrically or mathematically. We can talk about Rasch ability scores for they are scores that represent the amount of something that a person holds. The standard error of measurement is the statistic that tells you the precision for that type of score. Another class of scores that we deal with are scores that reflect standing in a distribution of people. These are the percentile ranks, standard scores, and the new NCEs. The index of precision for that type of score is the reliability coefficient. The standard error of measurement and reliability statistics are not interchangeable even though many people in the field assume that they are. When we talk about single scores, we talk about the standard error of measurement of a single score or the reliability of scores that come from that test. When we evaluate the precision of difference, then it is a standard error of measurement of the difference score that is the appropriate statistic, not the reliability of the differences. The reliability of the difference scores is used when the concern is whether there is a change in standings of a person. And so, in the field we frequently run onto statements to the effect that we shouldn't use a particular procedure when you're comparing two test scores because the reliability of the difference is low.

You mentioned the importance of having tests with adequate reliability, with a common criterion of adequate being .90 or higher for individual decisions. What if you don't have a test with reliability that good? You still need to make a decision. Don't you still want to make use of the best information that's available to you? The user, of course, has to appreciate the fact that there is much more chance imprecision in those data being used to form a decision. Also, it is inappropriate to compare reliability coefficients for tests when those reliability coefficients are taken from different studies. The reliability coefficient is a function of the standard error of measurement and the standard deviation of the sample to whom the test was administered. Thus, the coefficient will change depending upon the sample tested. One of the things that we need in the field is comparative reliability coefficients, where the same tests are given to the same subjects. The Institute might easily collect these in some of your planned research.

In regard to the planned studies, I have a few brief comments. I do not understand why the term "norm-referenced" is used in Research Area 1, where you are concerned with the adequacy of "norm-referenced" data for the prediction of success in a highly systematic instructional program. You could be using raw scores or other types of scores. It is the content of those tests that is important.

My comment on the study in Research Area 2, where you use a computer to look at technically adequate devices in assessment and decision making is related to my previous point: when you are asking people about tests that have sufficient reliability, maybe a more appropriate question would be "Which tests have the best reliability for a given purpose?"

Research Area 3 involves the comparative research on children who are failing academically and labeled LD with those who are not labeled LD. I would recommend using the independent Scholastic Aptitude clusters as the measure of ability rather than the Broad Cognitive Ability cluster of the Woodcock-Johnson Psycho-Educational Battery. Each of the Scholastic Aptitude clusters is a special "intelligence" test. Also, when you are measuring written expression, you might want to include the Proofing subtest as well as the Dictation subtest.

Follow-up Comments to Reactants

James E. Ysseldyke
University of Minnesota

I want to thank the people who have provided very insightful reactions to the paper. We'd rather have people tell us now those aspects of our research that we should be concerned about than to tell us five years from now. There are some overriding concerns and issues that we find ourselves dealing with in probably the same way that the reactants do.

I think all the reactants raised some issues regarding the language we use in the field of learning disabilities. Terminology is important. There is a tremendous variability from location to location in terms of the kinds of children we are talking about. The Chicago Institute has expressed their large concern with definitions for children in both suburban and inner city settings. But, it is a terribly important issue for us also.

Dick's comments regarding differentiating between interpretation and the norm-referenced/criterion-referenced testing is important for us to deal with. I think our approach agrees with your recommendations for interpreting data: any test is merely a sample of behavior.

Two things that I didn't want to miss responding to on Dick's concerns were the notion of the reliabilities that were recorded in the tables and the fact that the rates are there because they differ when they're standardized on different populations. The data that are reported in those tables are taken directly from the manuals and are the standardization data that the authors themselves provide. The ranges reflect simply the fact that reliability changes as a function of the age level of the individuals involved. I have trouble dealing with the notion that a .90 criterion might be unrealistic and that we should use the best available device. The best available device, in some cases, has a terribly low reliability. Decisions are made on the basis of unreliable data. I think that we ought to be very honest about what we are doing and say to the parents and teachers and to others that we just don't have a reliable and valid means of getting at a particular concern.

Barbara's comments on having a decision-making model to guide our efforts are extremely helpful. Although I agree, we are having trouble finding relevant models. The heuristics notions of Shavelson and Kahneman and others make sense but we have had some difficulty seeing how they make sense specifically in the placement team context. We have not been able, at this point, to identify a model to use to analyze what goes on. Instead, probably the model we are using to analyze what goes on is just straight-out naturalistic observation in the situation. At some point in the future, we will attempt to relate our observations to existing models.

Highlights of the Roundtable Discussion Following Ysseldyke Presentation

Discussion began with several Roundtable members expressing definitional concerns. Questions were asked regarding the nature of behaviors the Institute researchers would use as both marker variables, for the purpose of describing subjects, and as definitional variables, for the purpose of defining children as learning disabled. Alternative definitions were described and discussed, ranging from definitions based entirely on school labeling practices to test-based definitions. It was observed that much of the research being conducted by Institute personnel is not dependent on defining populations of children as learning disabled. However, the results of several Institute research studies should shed light on the important definitional problem.

The strategy being used by the five Institutes to deal with the definitional issue will be to describe the populations with whom they work on the basis of the same variables, such as age, grade, SES, and measures of ability, achievement, and interest. The Institutes will use the Woodcock-Johnson Psycho-Educational Battery to obtain the latter measures. In this way, identical descriptive data will be gathered and the results compared. In the long run, the Institutes will attempt to make statements of the relevance of specific research findings to a variety of settings.

The marker variable study being conducted by Barbara Keogh at UCLA also was cited as an important step toward clarifying the necessary components in descriptions of the research population. In its first year, the marker variable project conducted a comprehensive survey of the research literature and found inconsistent and non-comparable characteristics being used to describe learning disabled populations. It was noted that, given the dismal situation related to definition in past and current research, any attempt to determine whether youngsters in the research samples of the five Institutes are similar would be an enormous step.

Roundtable discussion continued with consideration of the issue of what to assess. Specifically, the distinction between the ability/process orientation to assessment and the skills orientation to assessment was questioned. It was noted that the two approaches have differed within special education. On one hand, the concern has been to look for difficulties within a child, in terms of information processing variables, psycholinguistic abilities, perceptual abilities, and so on. Within special education, this ability/process orientation has attributed the failure of a child to acquire academic skills to causes within the child. A contrasting approach denies the value of searching for within-child causes and turns to a more curriculum-based task analytic procedure. This skills orientation has sought to identify the child's skill level and then work from there to develop higher level skills in a sequential manner. While there has been considerable empirical support for the task-analytic skill-based approach to intervention, there has been little empirical support for the payoff from ability/process assessment and ability training intervention.

It was suggested that the distinction between the ability and skills approaches might be an historical one rather than a logical one. The two approaches cannot be distinguished on the basis of their focus on cognition (since the skills approach ultimately is tied to cognition theory in the way that it task-analyzes skills). Rather, one approach did not seem to work (the ability approach), so another approach that did seem to work (the skills approach) was adopted. This does not mean that the ability approach cannot work, but perhaps more realistically, that it has not been tested adequately. It was observed that individuals who have attempted to train processes have used skills as their dependent measures, but generally, have not taught these processes (e.g., perceptual-motor processes) in the specific domain to be tested (e.g., reading). Teaching processes within specific domains might lead to different results than previously reported.

Another issue raised was related to the suggestion that decisions should be deferred when only technically inadequate information was available to a decision-making team. It was argued that, given the current functioning of multi-disciplinary teams and the different levels of decision making, decisions cannot be held in abeyance until adequate test instruments are available.

It was agreed that we are forced to make decisions; we cannot call for a moratorium on decision making in educational settings. However, the problem with using the "best" available tests in most cases is that we cannot demonstrate that the measures we select are indeed measuring what they purport to measure. One solution is to have alternative means of data collection which are not test-based, particularly for making instructional intervention decisions. It was noted that the best way to determine what content to present a child and what instructional methodology to use is to teach the child.

Decision-making teams, however, must make decisions related to other than instructional concerns. They must make decisions regarding placement, related services needed, parent involvement, and so on. The question of what sources of information should be used to make these kinds of decisions was raised.

It was suggested that placement decisions need to be test-based decisions, but that documentation of the need for a specific placement must be made as well. Considerable data should be gathered on the alternative instructional techniques and strategies that have been attempted in a youngster's educational program, and whether they succeeded or failed. In some cases, decision makers will simply have to admit that they do not have adequate means for making a decision.

The interactive nature of assessment and intervention was then discussed. It was suggested that assessment which results in identification and classification clearly has an effect on instruction. Special education training programs provide different courses in curriculum and in methods for children given different labels, such as mentally retarded, emotionally disturbed, learning disabled, autistic, physically handicapped, and so on. Although not necessarily an approach to be advocated, the classification procedure purports to tell us how to teach.

However, the claim that assessment and teaching are interactive implies a two-way street. Behavior analysis is an excellent example of the traffic flow where assessment influences intervention. But, what evidence is there that the traffic flows on the other side of the street, that intervention influences assessment? Specifically, how does what happens to a child in the classroom influence how that child gets tested?

Several responses were made to this question. At a gross level, a teacher referral indicates that the child's educational program has in some sense been inadequate; this event sets into motion a series of assessment activities. Those who are working with children ought to gather data on the intervention process. The interventions tried should dictate the data to be collected. Intervention and assessment would then go hand-in-hand.

The concern was raised that while continuous gathering of data may occur prior to a decision, once a child is placed, the decision seems to be chiseled in stone. Data collection is not continued once a child is placed. Assessment has not received the emphasis it should have in re-evaluating placement decisions. Once made, placement decisions should be considered as tentative, subject to further input from other data gatherers.

It was suggested that while the idea of rethinking placement decisions is important, the present team decision-making process seems to inhibit its occurrence. The placement process is a cumbersome one which overburdens the individuals involved. However, the system does have a built in annual review requirement, with re-evaluation required every other year. The latter requirement should force educators to look at the appropriateness of the continued placement or categorization of the youngster. Unfortunately, the system is difficult to work within and typically is very slow moving, even when there is information which dictates that another decision should be made.

It was noted that data collected through standardized tests somehow are valued more highly in decision making than data collected through direct observation by teachers in a more informal and unstandardized way. This happens even though the unstandardized, direct observation data could be fed back into the decision-making process more regularly. Such biases may be a reason for the apparent greater influence of a school psychologist's opinion in decision making than the opinion of the close, day-to-day worker with the child.

One aspect of the Institute's research is specifically concerned with such interactions and influences in placement teams. Researchers also will be studying the extent to which intervention data actually are input into the decision-making process. Although intervention data ought to be important in making decisions about placement, the extent to which this actually occurs is questionable.

The issue was then raised as to where the Institute's research on placement team decision making would lead. It was suggested that the findings should somehow lead to an improvement in the decision-making process. However, there is a problem in that there does not seem to be a dependent variable: we do not know when a decision is right and when it is wrong.

It was suggested that the dependent variable data will probably have to come from pupil outcomes. How the pupil does will determine whether a decision was right or wrong. The result of the research should have great potential for influencing the process. At the present time, the research concern is to describe what happens. Such a description should lead to some statements about desirable decision making. In the future, the Institute will probably be engaged in specific research to influence the decision-making process. One approach would be to compare and contrast alternative decision-making models. Another approach would be to train placement teams in decision making and look at the influence of such training. The present descriptive study might indicate that people spend 90 percent of their time in decision-making meetings presenting data. Given this finding, it might be possible to predict with 90 percent accuracy what the decision would be if the data were available ahead of time. Thus, a third approach might be to enter all data into a computer, give the decision-making team the computer-based decision, and then evaluate the influence of this information on the team's decision. While there are a great number of possibilities for influencing the decision-making process, empirical evidence is needed first on what currently happens in the entire assessment and decision-making process.

Highlights of the Open Discussion Following Ysseldyke Presentation

The Open discussion began with comments on the research to be conducted at the Institute. It was noted that the direction of the research seems to be to identify some general principles, ones that can be used effectively in applied settings. It was suggested also that despite their commendable emphasis on the non-testing aspects of the assessment process, the researchers should realize that observational methodologies, as well as interviewing, are not simple cure-alls. Many problems will be encountered in attempting to use an ecological approach. Another comment dealt with the notion of the multidisciplinary team. Although called "multidisciplinary," team members really are very much alike. One might speculate that if other individuals were included, individuals who traditionally have not been involved in the psychoeducational system, the decision-making process might be very different.

The possibility of considering the weight that a person assigns to a given piece of information when making educational decisions was then raised. For example, a person might be given a great deal of empirical data yet reject it because he or she has a mental model established about what is important in identifying a learning disability. Will there be any attempt to determine the weight of a person's mental model and the extent of empirical data needed to override the model?

In response, it was agreed that one's model or set does influence decision making. The simulated decision-making research will address this issue to some extent. It will allow the researchers to introduce certain data (such as sex, SES, and attractiveness information), analyze their effect on decision making, and then compare these results to data obtained in studies in which decision makers report the extent to which such information influences their decisions. It might be predicted that decision makers say the information does not influence their decisions even though research findings typically indicate that they do. Research on attractiveness bias, for example, has indicated that a child's looks do not just create an initial impression, but rather influence grading and other educational events that occur throughout the school year. It was suggested that because such internal influences clearly exist, it is important to build them into the system to be recognized and studied.

The point was then made that such simulated bogus study information can have restricted value because only one variable is studied at a time. The decision makers are given information that varies on just one variable. It was suggested that if a decision maker is presented all data, then a factor that otherwise might appear to have a significant effect would be washed out.

In response to this, it was noted that the research did not study just one variable. Even in practical situations (the classroom), where all variables were operating, a variable like attractiveness still was operating. It did not wash out. It was suggested, however, that new information does serve to modify one's previous model. Weights are put onto new bits of information and these weights determine the influence on one's model.

Another audience member suggested that some very important factors have been ignored in research on decision making. These are factors that exist out in the real world of education - declining enrollment, fear of loss of jobs, and other irrational variables that do enter into decision making. It was agreed that such variables certainly are important and that the research will attempt to look at them.

The question was next raised regarding the extent to which the Institute would be investigating the re-evaluation process. Once a child is placed, are needed changes perceived, and if perceived, do they actually occur? It was noted that although such investigations are not explicitly included in the proposed research, they certainly will be an area of concern for subsequent research. Studies would involve determining the extent to which data on a youngster's success, in a variety of interventions, are used in the process of making placement or continuation of placement decisions.

The discussion concluded by shifting back to the definitional issue. It was suggested that the treatment of a learning disability might be greatly improved if its etiology were known. The situation might be considered similar to that in the areas of suicide and fevers. Suicide is a behavioral symptom. Initial research which attempted to correlate various factors with the occurrence of suicide was not very helpful in preventing its occurrence. The etiology makes a big difference in treatment. Similarly, fever is a behavioral symptom; but, treatment relies on knowledge of etiology. Learning might be considered in the same way, as a behavioral symptom that needs to be treated with consideration being given to etiology. It was argued, with some disagreement, that whereas the dependent measure in suicide or fever is quite reliable, the dependent measure in the identification of learning disabilities is not quite so reliable.

In relation to the definitional issue, it was noted that the Institute plans to look at learning disabled students and students who are failing academically but not labeled "learning disabled." Data on these children will be collected, then presented to diagnosticians to see whether they can identify those children already labeled as learning disabled. If this research indicates that diagnosticians cannot agree, the field really has problems. Furthermore, researchers are faced with a dilemma, for it will be impossible to reliably identify who should be considered as learning disabled.

Behavioral Research Methodology as a Basis for the Formative Evaluation of Learning Disability Service

Stanley L. Deno
University of Minnesota

In analyzing the contributions that the behavioral approach might make to the assessment of children with learning disabilities, three general categories of assessment decisions were considered. First, identification decisions (which include referral, screening, and classification); second, program planning decisions (which include goal setting, curriculum and instruction, and level of service); and finally, program effectiveness (which includes both formative and summative decisions). The two types of decisions in the program effectiveness set require some explanation. Formative decisions are the decisions made by people during the course of implementing program plans. Formative decision making requires answers to questions such as: "Do the methods and materials seem to be working?"; "Is the level of service appropriate?"; "Are the parents satisfied?"; "Should we change something?" Formative decisions are made, then, to help improve and adjust a program on a continuous basis. Summative decisions are an outcome of judgments made either annually or upon program completion. Summative decisions are after-the-fact decisions typically made by those persons empowered to certify that a program succeeded. The essential summative question is: "Did the program succeed?" An implied question is: "Should we allocate special education resources similarly in developing future programs?"

While the Institute for Research on Learning Disabilities has taken on the task of understanding and improving assessment practices for children considered learning disabled, this project's goal is to bring the behavior analysis perspective to bear on the problem of assessment for purposes of making those three general sets of decisions. In doing so, we have identified a set of relevant "perspectives,"[1] a few of which I am going to share with you here.

First Perspective

> The assessment of problem behavior should be approached from the point of view that it evolves from and is governed by the same laws that govern the development of normal behavior.

This first perspective deals with the assessment of problem behavior, the kind of behavior that children must emit to be given a special label, such as "learning disabled." The perspective implies that when a child's behavior differs significantly enough for that child to be referred, one should approach assessment as if the behavior to be assessed can be accounted for without invoking a special set of determinants (variables) to explain the occurrence of the problem behavior. The behavioral perspective is that problem behavior, as well as normal behavior, is a function of three sets of variables: (1) the genetic constitution of the individual, (2) the training or reinforcement history of the individual, and (3) the current environmental circumstances of the individual. Often, people assume that behavioral psychologists think of the human organism as beginning with a tabula rasa; in other words, that all human organisms are the same, without any constitutional variation. I do not think that assumption is representative for most behavioral psychologists. From my own point of view, it is just as useful to assume that there may be genetic constitutional differences, but such differences are probably the proper domain of physiologists and perhaps physiological psychologists. Behavioral psychologists typically do not study the influence of an individual's genetic constitution. Instead, they focus on the historical and current interactions between individuals and their environments in an attempt to account for changes in behavior.

---

[1] These are presented in Deno, S., Mirkin, P.K., & Shinn, M. Behavioral Perspectives on the Assessment of Learning Disabled Children. In J. E. Ysseldyke and F. Morrison (Eds.), *Multiple perspectives on assessment of learning disabled children*, in press.

Just as people sometimes erroneously conclude that behavioral psychologists do not believe in constitutional differences, I think people also sometimes assume that behavioral psychologists approach the analysis of behavior as if there is no history -- that all behavioral determinants can be found in the current environment. I doubt seriously that any behavioral psychologist would say that. Generally, the assumption is that each individual carries with him or her a mass of experience which is described in terms of the reinforcement contingencies experienced by the individual. That reinforcement history combined with constitutional variables, then, determines how the individual is likely to respond in the current environment.

The interplay among variables of the genetic constitution, reinforcement history, and current environment was succinctly characterized by Bijou (1977) in an article that he published in Exceptional Children a couple of years ago. He called the model "interactional," saying it consisted of two basic concepts. The first concept is that there is a continuous interaction between biologically developing children and the progressive changes in their environments. The second is that these interactions change the individual (that is, develop a person with a unique personality) on the one hand, and change the environment on the other. The individual acts on the environment and is in return influenced by the changes that his or her actions bring about in the environment.

At this point it should be emphasized that while the behavioral model, its principles, and its procedures can significantly contribute to improving learning disablities services, the focus of the present research is not on the application of behavioral principles to learning disabilities. Instead, the focus is on the research methodology used by behavioral psychologists to empirically analyze the behavioral effects of changes in the individual's current environment. Reinforcement, punishment, stimulus control, and schedules of reinforcement, are principles that ought to be applied in developing learning disabilities programs. In fact, as I train teachers, I require them to use those principles in the development of interventions. However, in approaching the problem of assessing children with learning disabilities, we find more relevant the scientific methodology that behavioral psychologists have developed to experimentally analyze interactions between behavior and the environment. Thus, the application of behavioral research methodology to assessment of students called learning disabled is the direction of the present research.

Second Perspective

> The methodology which has been developed to accomplish behavioral assessment is characterized by the careful description of the behaviors of interest, the development of procedures for directly observing and recording that behavior over time, and the use of time series research designs to try to determine functional relationships between changes in the environment and changes in individual behavior.

The methodology is idiographic rather than nomothetic, and it may be developed to make unique contributions to the assessment of LD students.

Time series research designs are not the private domain of behavioral psychologists. One of my favorite illustrations of the application of time series data analysis to look at human behavior comes from a Common Cause newsletter. The particular time series datum collected was the number of pieces of franked mail that our legislators in Washington send home to their constituents over time. The question was, how much and when is franked mail used? The data are presented in Figure 1.

Insert Figure 1 about here

As can easily be seen by inspecting the graph, a relationship apparently exists between when elections occur and how much franked mail is distributed. What you would observe if this were a cumulative graph is the characteristic fixed interval scallop. An increase in use of franked mailings occurs up to the point that the elections are held, and then an immediate drop occurs in how much free mailings legislators use. The amount begins to increase again over time as an election approaches, and then it drops off.

One of the nice things about time series data analysis is that it can be used to keep our legislators accountable. More importantly for us, however, is that it can be used in ongoing natural experiments. Careful presentations of time series experiments have been made by Glass, Willson, & Gottman, (1975) and by Hersen and Barlow (1976); many good articles exist on time series analysis. An important feature of Glass, Willson, and Gottman's book, however, is that they point out the potential contribution that time series research design can make to the evaluation of educational programs on a continuous basis. If you know what data might be useful and significant, you can then routinely collect those data, identify naturally-occurring changes in the environment, and then determine what relationships might exist between changes in the data and changes in the environment.

Probably all of you have seen graphs like that in Figure 2. Most of the time they appear in technical reports written by behavioral psychologists. Figure 2, however, presents a graph that a teacher developed to monitor the percentage of a class which was engaged in "disruptive noise."

- - - - - - - - - - - - - - - - - - -

Insert Figure 2 about here

- - - - - - - - - - - - - - - - - - -

This graph shows variation in percentages of disruptive noise across time, with vertical lines denoting points at which specific environmental changes occurred. The graph also depicts a "reversal design," in which an attempt is made to replicate the effect on behavior of changing from baseline to the treatment condition. Replication, of course, is not always possible, but it is a way to eliminate plausible rival hypotheses for explaining the behavior changes. In this example, there is a decrease in the behavior both times the treatment is introduced. Many graphs of this type exist throughout the literature of applied behavior analysis. I have always been intrigued not only by the treatments but also by the potential for evaluating individual special education programs which is inherent in the research methodology. Special education programs are individually-oriented intervention programs. When looking at individuals instead of groups, a methodology is needed to help one look at individuals to determine what influence a program is having on them. Conventional research methodology typically contrasts performance between groups rather than within individuals. When approaching the problem of assessment within learning disabilities programs, then, the within-subject approach of behavioral psychology seems more useful for assessing treatment effects with individual learning disabled students. A related perspective is relevant at this point.

Third Perspective

> Behavioral assessment requires first, a precise description of the level and direction of the behavior, and second, an empirical analysis of the variables in the current environment which control the level and direction of that behavior.

Unlike most approaches to assessment, the relevant data in the behavioral approach are not simply measures of the current level of performance (or achievement quantified as raw scores, grade equivalent scores, or percentile scores). Just as important are data on the direction in which a behavior is changing (i.e., is it increasing or decreasing?), since the behavior may not be static. While it is possible to determine level by assessing a behavior at a single moment in time, the behavior may, in fact, be increasing. John Stephens (1967), in his book entitled The Process of Schooling, emphasizes the fact that children are getting better in basic skills all the time, and argues that the schools do very little to influence that in any specific way. He hypothesizes a kind of general press that the school brings upon the child, to learn to read, write, spell, and so forth.

Stephens' point that academic growth is continuous implies that assessment of a child's behavior requires description not only of current level of performance at a moment in time, but also the trend or rate of change in that performance over time. Rate of change is essential, especially if one is going to use assessment data to evaluate the effects of services. If we agree with Stephens that children's academic performance is not stationary, observed changes in a student's achievement might well be attributable to conditions other than those arranged by the special educators.

Nowadays, special education is charged with demonstrating its efficacy to the consumer. Pre and posttesting is a common practice to provide evidence of program effectiveness. It should now be obvious that pre and posttesting is an inadequate approach to evaluating treatments. Here is an illustration from a class I teach.

- - - - - - - - - - - - - - - - - -

Insert Figure 3 about here

- - - - - - - - - - - - - - - - - -

Assume that the little squiggles on each graph in Figure 3 represent individual academic performance data, and that the heavy vertical lines in each case represent instructional intervention. It is possible that the effect of each intervention will be interpreted quite differently depending on whether one looks at level of performance alone or whether one considers level and direction together. To illustrate the problems of pre and posttesting to evaluate intervention effects, I propose to my students that they interpret intervention effects first without the time series data and then with the time series data. I ask them to assume that the pretest datum point is that point immediately preceding the intervention (which we ordinarily obtain during initial assessment), that the intervention line represents the onset of learning disabilities service, and that our posttest datum is the last point on the graph (e.g., a year-end assessment). Such a situation probably characterizes much of the assessment that goes on in evaluating services. One assesses where the student is, then sets up a program and teaches for a while, and finally, determines whether the student has increased his or her performance on a set of objectives, or has mastered a certain set of skills. If student performance changed from point A to point B, it is assumed that the intervention program produced the effect. As inspection of the time series data in each graph illustrates, however, it is quite possible the conclusion will be different if one has available more information. In several instances, the growth from point A to point B is perfectly predictable over a period of time based on the actual data from baseline. In those cases, no evidence is present that the intervention had any effect whatsoever. In other cases, while a "no effect" conclusion would be drawn from pre and posttesting, the complete time series would indicate otherwise. In the most embarrassing cases, we might assume that our intervention has had a positive effect on the student's behavior when inspection of the time series reveals that onset of learning disabilities services actually depressed or decelerated the performance.

As a result of this exercise, the point is usually quite obvious to the teacher in training. If one really wants to be sure of the effects of instruction, one needs to assess performance in such a way that estimates of trend, slope, or direction, as well as level, may be obtained. That, I believe, is an important characteristic of the behavioral approach to assessment, especially in contrast to other approaches to assessment.

Fourth Perspective

> The comprehensive behavioral teaching systems used to design remedial instructional programs differ significantly with respect to assessment.

Several well-developed behavioral teaching systems exist which can be adopted and used extensively in school programs. In the review of the data base, it was noted that those applied behavioral systems differ with respect to the emphasis that they place on the collection of the kind of time series data that I've been discussing. Some of those applied behavioral systems tend to rely more on pretesting and posttesting and may use post hoc teacher judgment on checklists as the primary assessment datum for deciding whether or not a student actually has a set of skills. Other applied behavioral systems heavily emphasize daily data collection. A fair illustration of those differences exists in the Distar system which tends to do frequent pre and posttesting, and in Precision Teaching where daily measurement is the essence of the system. Owen White (1977) has described the Distar developers as interested in specific prescriptive solutions to teaching particular skills. In such an approach, if the teacher has a student who is deficient in a skill, that teacher is given a teaching package which is used to teach the child that skill. In contrast, people in the Precision Teaching tradition are given a package for creating time series data on the specific skill, and the teacher is trained to use the daily time series data to evaluate the effectiveness of instructional interventions, packaged or not. Since many variations in instruction are possible, the teacher is given the tools to try to appraise the effects of different alternatives.

Fifth Perspective

> The applied behavioral systems which contribute most uniquely to the development of procedures for assessing children with learning disabilities include repeated behavior sampling and graphic displays of time series data.

We believe that the behavioral systems which might make the most unique contribution (not necessarily the greatest) to assessing children with learning disabilities and to educational assessment generally, are those which use or require repeated behavior sampling and graphic displays of time series data. This is not to say that what Engelmann and others who apply behavioral principles to develop prescriptive techniques are doing is wrong, or not useful. Data exist quite to the contrary. The point is that when it comes to assessment methodology, the use of daily or routine repeated behavior sampling and visual display of data, might make a unique contribution.

Since the potential of time series data in assessment is our primary interest, the important distinction for us in testing is not whether assessment is norm-referenced or criterion-referenced; rather it is whether assessment involves single or repeated behavior sampling. In this regard, Wells Hively has made somewhat the same point that was made earlier today -- criterion-referenced tests and norm-referenced tests can be used for the same purposes. One can convert most norm-referenced tests into criterion-referenced tests, and vice-versa.

We take then, from behavioral psychology, repeated measurement and time series data analysis; it is around that approach to assessment that we have developed a program of research. And, while a review and synthesis of the area suggested the potential contributions of behavioral research methodology to identification and planning decisions, we have restricted the focus in our research plan. The decision made was to focus on assessment for purposes of making formative evaluation decisions -- the daily decisions made by teachers in an attempt to improve a program which is in progress. Thus, the primary focus is on assessment that can be accomplished by the teacher when instructing, or managing interventions, and which can be used to tailor the instructional program to the individual student in an effort to make that instructional program more successful. At this time, the research does not include using a behavioral approach to identify children with learning disabilities, or to prescriptively plan programs (although we will be involved in program planning). We expect, however, that the research will have implications for all phases of programming. We simply will not work on all fronts simultaneously. At some later time, perhaps, we will have the opportunity to try to do that.

The Research Proposal

To this point we have considered general perspectives which have provided the background for our research. We turn now, more specifically, to the research, beginning with the assumptions upon which it rests.

Assumptions

The first assumption, which I believe very strongly, is:

> At the present time we are unable to prescribe specific changes in the programs of individual students that will certainly be effective; therefore, changes in a student's instructional program should be treated as hypotheses, which must be empirically tested.

If you do group research, you can identify interventions which might work in general or might work in a high proportion of cases. Special education, however, is required to do individual assessment and programming. That creates an incredible technical problem. To predict that a particular intervention, even a placement in a resource program or in a special class, will be an appropriate change for a given child is analogous, I think, to requiring the National Safety Council to predict not only how many accidents there will be on the fourth of July, but who will have those accidents. I could be wrong, but I think that our problems in assessment and programming are analogous to that. In individual cases, it is very difficult to predict with any certainty whether any kind of reform that we make in the student's program is going to benefit that individual student. If you believe that people are unique individuals interacting uniquely with their current environment as a function of their unique training histories, and their unique constitutional characteristics, then the difficulties in attempting to make a reliable prediction about what program, what curriculum, what incentives, managed by what teacher, obviously exceed our technical competence. A very highly structured program seems to work for a lot of children; but, to predict exactly which kids that program is going to work for is virtually impossible. There is always that residue of students who do not succeed very well in even our best programs. Further, those kids who are succeeding may be capable of greater success given other program arrangements. Our view, then, is you cannot be sure about what is best for the student on an a priori basis; you must continually monitor the performance to ensure that what you are doing is benefiting the child. That's the ideal system.

Our second assumption is that:

> Special education is an intervention system created to produce reforms in the educational programs of selected individuals.

One of my favorite articles is one by Campbell (1969) appearing in the American Psychologist, entitled "Reforms as Experiments." His point there, as mine has been, was that we should treat all administrative and educational reforms as experiments, doing the best possible experimental analysis of the effects of those reforms that we can. Special education for each child is an intervention or a reform which can, and now with due process requirements, must be empirically tested.

From our primary assumptions it follows that:

> The effectiveness of learning disability services provided for an individual student will be significantly

> determined by the capacity of those services for continuously assessing the effects of program adjustments which have been designed to improve program effectiveness.

This assumption is critical to the proposed research. The assumption implies, in a statistical sense, that of the total variation in student achievement, a significant proportion is going to be accounted for by the daily instructional decisions made by the teacher. The assumption is that what the teacher does in making adjustments in a student's program has a significant impact on the success of the program.

A final assumption is that:

> Teachers will have to be capable of implementing and managing systematic formative evaluation which makes a program responsive to its effects with individual students.

If it is true that the effect of teacher decision making is significant, then teachers must be able to continually assess the effectiveness of what they are doing. This implies that if we can create better formative evaluation systems and effectively teach teachers to use those systems, they are going to be more successful. I believe that the really important things that happen to students are what teachers do from day to day, and that we can positively influence what they do by improving the assessment procedures teachers use in making these routine decisions about reforming a student's program.

I believe that systematic formative evaluation systems will have to be identified or developed which can test the effectiveness of adjustments in a wide variety of program variables. Consider first, changes in level of service as a variable. Our evaluation systems should help us to determine whether or not changes on some continuum of services from Level I to

Level II, to Level III, and so on, are benefiting the child. The evaluation system should also assess effects of changes within a level: Suppose a student is receiving thirty minutes of LD resource assistance, and the time is then increased to forty-five minutes or an hour. It is desirable to be able to appraise the value of that program reform occurring within one level of service. The formative evaluation systems should allow teachers and other decision makers to determine whether or not such an increase in amount of service has an effect. Changing the amount of time in instruction is important since it involves additional resources. At the same time many believe that amount of instructional time is a significant variable in determining instructional success.

Our formative evaluation system should also allow determination of whether or not the changes between curricula, such as those between different reading systems, are significant. Further, since changes occur within a curriculum, such as skipping steps or modifying the skill sequence, it is important to attempt to objectively assess whether or not changes within a curriculum are benefiting the child.

A third set of variables to consider is changes in the type of instruction or intervention. Introducing an explicit contingency management system, providing oral reading practice, or providing auditory discrimination training, are attempts at program reform which may or may not be benefiting the child. As with changes in level of service and in curriculum, the number of changes in instruction or intervention which can be created is almost endless, and their effects are unpredictable. To compound the problem, even when a program appears to be working, it may be necessary to change anyway (cf. White & Liberty, 1979).

Conclusions

The assumptions specified above and the perspectives from the behavioral approach to assessment can be combined into a set of conclusions about research which must be done.

First, time series research designs based on repeated measurements of student performance, which behavior analysts have developed, could be used to systematically test program adjustments. It ought to be possible to evaluate the effects of changes in learning disability services by using the kind of data collection and display systems that have appeared for years in the *Journal of Applied Behavior Analysis*. Often, when I make that suggestion, too many people react by saying, "Hey, that's behavioral; what they are looking at is the effects of reinforcement contingencies on the behavior of the students. The children I'm interested in are learning disabled. That behavioral stuff does not really apply because the child's problem is not a reinforcement problem, it's a learning disability problem." So the research is set aside without considering its potential. I am saying, "Hey, wait a minute, you're throwing out the baby with the bath!" Whatever your theoretical biases, there is a really good analytic system involved, one which might benefit the children about whom you are concerned. You can use the behavior analytic research methodology without embracing either behaviorism or operant conditioning. Individual subject time series system analysis can help you to evaluate, in a formative way, the effectiveness of what you are doing.

A second conclusion is that, to be effectively used by teachers, the time series research designs will have to be embedded in formative evaluation procedures which can easily be used within the context of an ongoing learning disabilities program. One of the major resistances to repeatedly measuring student performance is that it is such an intrusive activity. Teachers question whether the time and energy devoted to it is warranted. The question typically asked is, "If I spend all my time testing, when am I going to teach?" We plan to address this concern in our program of research.

A third conclusion directing the research program is that _formative evaluation systems should include specification of target behaviors which validly index improvement in the curriculum domains for which learning disability services are commonly provided_. We should create simple and direct procedures for repeatedly measuring target behaviors, for graphic display of repeated measurement data, and for using the graphed data systematically to cumulatively improve a program. I am proposing that what we should do is develop procedures that are analogous, I suppose, to what physicians use when they measure vital signs. We need to know what are the vital signs of growth in reading and other basic skills which are easy to measure and can be used by teachers to monitor the effectiveness of what they are doing. A first part of the research program, then, addresses itself to the identification of target performance which validly indexes improvement in the curriculum domains for which learning disability services are commonly provided.

Research Objectives

Our primary three-year research objective is to empirically determine whether teachers who are using the kinds of formative evaluation systems that I have been describing are more effective in improving the basic academic skills and the social functioning of learning disabled children than teachers who do not employ such systematic formative evaluation. We are trying to achieve that in three stages.

The first of these, Stage 1 in our general research plan, seeks to answer the following question: Can a simple set of behaviors be identified in reading, spelling, written discourse, and social adjustment which validly index improvement in those domains? Some other basic skill areas have been left out deliberately, not because they're unimportant, but because we have only so much time to devote to this activity at the present.

Secondly, if we can identify behaviors as validly indexing improvement in those domains, can simple measurement procedures be developed for those behaviors which teachers could use routinely, daily if necessary, to monitor the effectiveness of interventions designed to change those behaviors? It is one thing, of course, to identify a behavior as a good index of growth, and it is another problem to develop simple procedures for routinely obtaining data on that behavior.

Third, does variation in the frequency or type of measurement affect outcomes? That may seem like a peculiar question, but Tom Lovitt has talked a lot about the value of direct and daily measurement. For purposes of developing formative evaluation procedures to evaluate the effects of the kinds of changes that we are talking about in curriculum, instruction, and level of service, we would like to establish the importance of daily measurement. We would like to know whether the teacher can drop back to less frequent measurement such as three times a week, or weekly, or once every two weeks, or once a month. We also seek to discover whether it makes a difference if behavior is quantified as percentage, or rate, and whether the duration of a measurement sample is one minute, three minutes, or five minutes. Variables like those, which are part of measurement itself, can conceivably influence student performance and teacher decision making, so they should be examined empirically.

An additional set of research questions relates to whether systematic techniques can be identified which improve teachers' use of daily assessment data. A few studies have been done in which specific rules have governed teachers' use of data to decide when a program should be changed. Ralph Bohannon and Phyllis Mirkin both used such systems in their doctoral research and obtained positive effects. Kathleen Liberty and Owen White are the leaders in this particular area of research. Our question is, if you have teachers using a data utilization system, does that maximize the potential contribution of time series data to improving a student's program?

A final Stage 1 question is, does variation in who does the measurement influence effective data utilization? Teachers, students, or parents can be involved in data utilization. Does it matter who is involved?

We turn now to Stage 2 of our research. Once we have answered the questions posed in Stage 1, we will attempt to construct formative evaluation procedures in each domain which are logistically feasible. I thank Tom Lovitt for the notion that we need to consider logistical feasibility within the context of school learning disability programs. To do so, we will attempt to determine whether or not the procedures we develop are generalizable across different curriculum methods, classroom organizations, age and grade levels -- whether it is easy to train people to use the procedures, whether the procedures are efficient with respect to teacher and student time, and finally whether it is easy for teachers, and others involved, to manage such systems. Those are critical variables in determining the likelihood that such systems would be adopted and used within the schools. Logistical concerns are major ones for answering our primary research question. Teachers must be able to use our formative evaluation procedures over the period of a full school year. If the procedures are too cumbersome, we will have difficulty maintaining teachers' measurement behavior.

If we are successful in Stage 1 and 2, we would like to do a major experimental contrast between the effectiveness of teachers' using the formative evaluation systems developed in Stages 1 and 2, and teachers' not using such formative evaluation systems.

I will close by providing a hypothetical example of what we hope to achieve. Consider the graph in Figure 4. Suppose, for example, it turned out in Stage 1 that reading common words in isolation from some high frequency word list, like the Dale list of 769 easy words, is a valid index of the student's proficiency in reading. Suppose further that we could generate a display of the student's performance using simple daily measurement procedures, and that what we displayed in Figure 4 is one student's performance from initial assessment through a series of program reforms. The numerals on the graph represent reading from the isolated word list during each program phase -- the rate correct and incorrect, median rates, for those phases. The first reform (placement decision) was to have a learning disability resource teacher consult with the teacher about some possible curriculum change. The changes, sequentially, were consultation with a curriculum change, but continued class placement, and then a second curriculum change.

Insert Figure 4 about here

As you can see, these reforms did not produce an improvement in reading performance (at least as indexed by our measure), so then we changed the level of service to resource assistance with one-to-one instruction for a period of time during the day. In doing so, we maintained the second curriculum change made in the classroom. The data indicate that the reform initiated growth in reading, but a bit later the growth seemed to level off. What we did then was to introduce an instructional change -- oral re-reading to mastery; again performance accelerated. This hypothetical example well illustrates, I think, what we are striving to achieve in our research. A system like this is the kind of system that we hope to develop and test in the final stage of our research.

References

Bijou, S. Practical implications of an interactional model of child development. Exceptional Children, 1977, 44 6-15.

Campbell, D. T. Reforms as experiments. American Psychologist, 1969, 24, 409-429.

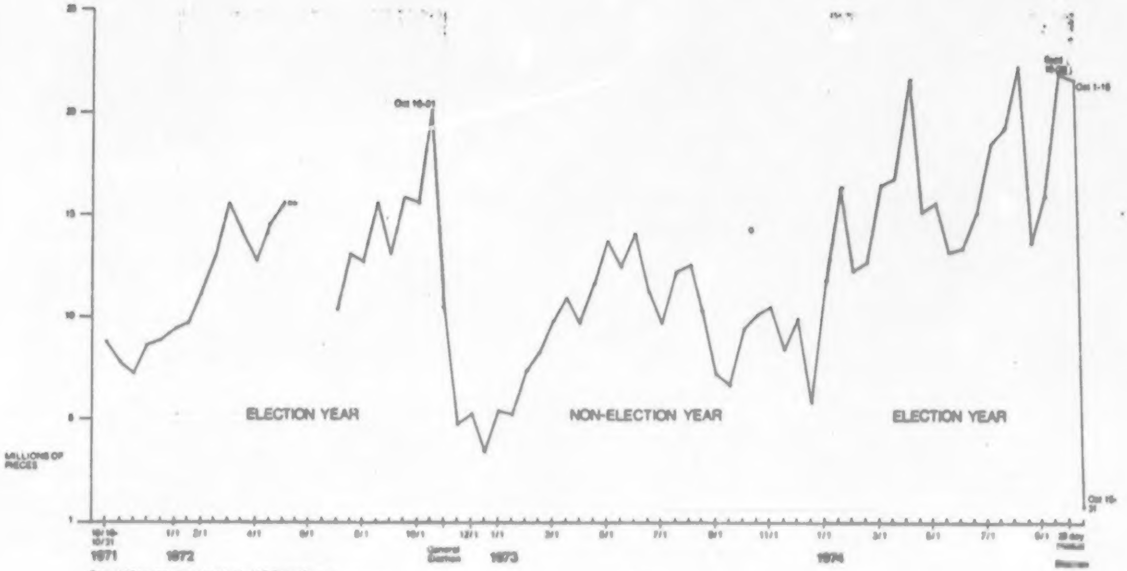
Glass, G. V., Willson, L. L., & Gottman, J. M. Design and analysis of time-series experiments. Boulder, Colo.: Laboratory of Educational Research, University of Colorado, 1975.

Hersen, M., & Barlow, D. H. Single case experimental designs: Strategies for studying behavior change. New York: Pergammon, 1976.

Stephens, J. M. The process of schooling. New York: Holt, Rinehart, & Winston, 1967.

White, O. R. Behaviorism in special education: An arena for debate. In R. Kneedler & S. Tarver (Eds.), Changing perspectives in special education. Columbus, Ohio: Charles E. Merrill, 1977.

White, O. R., & Liberty, K. Instructional hierarchies research project: Annual report for 1977-78. Seattle: University of Washington, 1979.



Prepared from documents provided by U.S. Postal Service.
**The figures for this period were not available when chart was prepared.

Figure 1

Use of Franked Mail by U.S. Congressmen during Two Election and One Nonelection Years*

*Source: Common Cause Report, 1975, 5(7), 8.

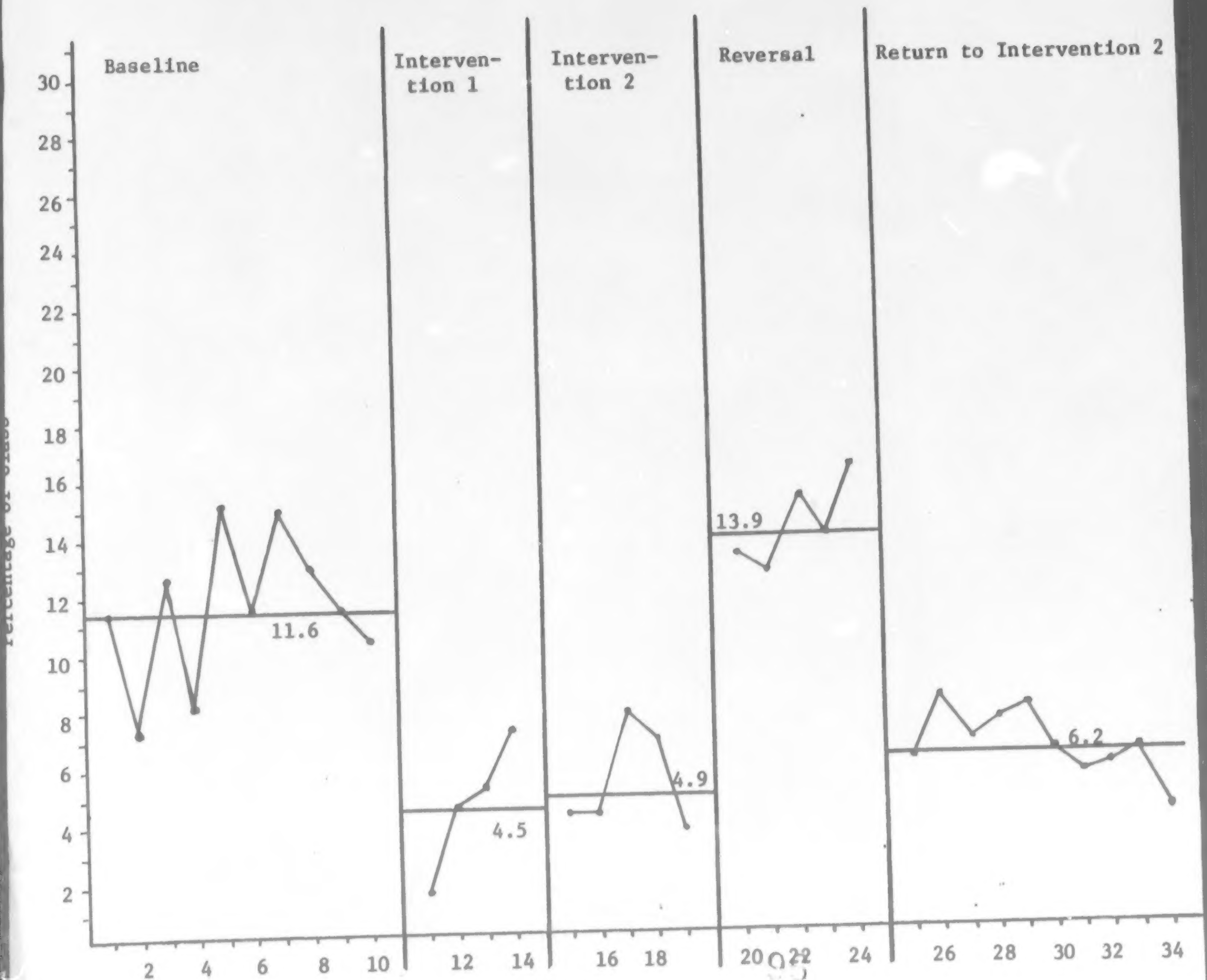


Figure 2

Percentage of Class Engaged in Disruptive Noise

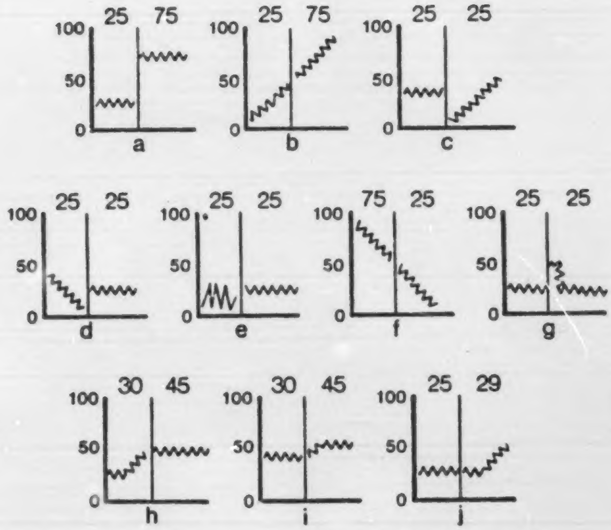


Figure 3

Examples of Performance Data

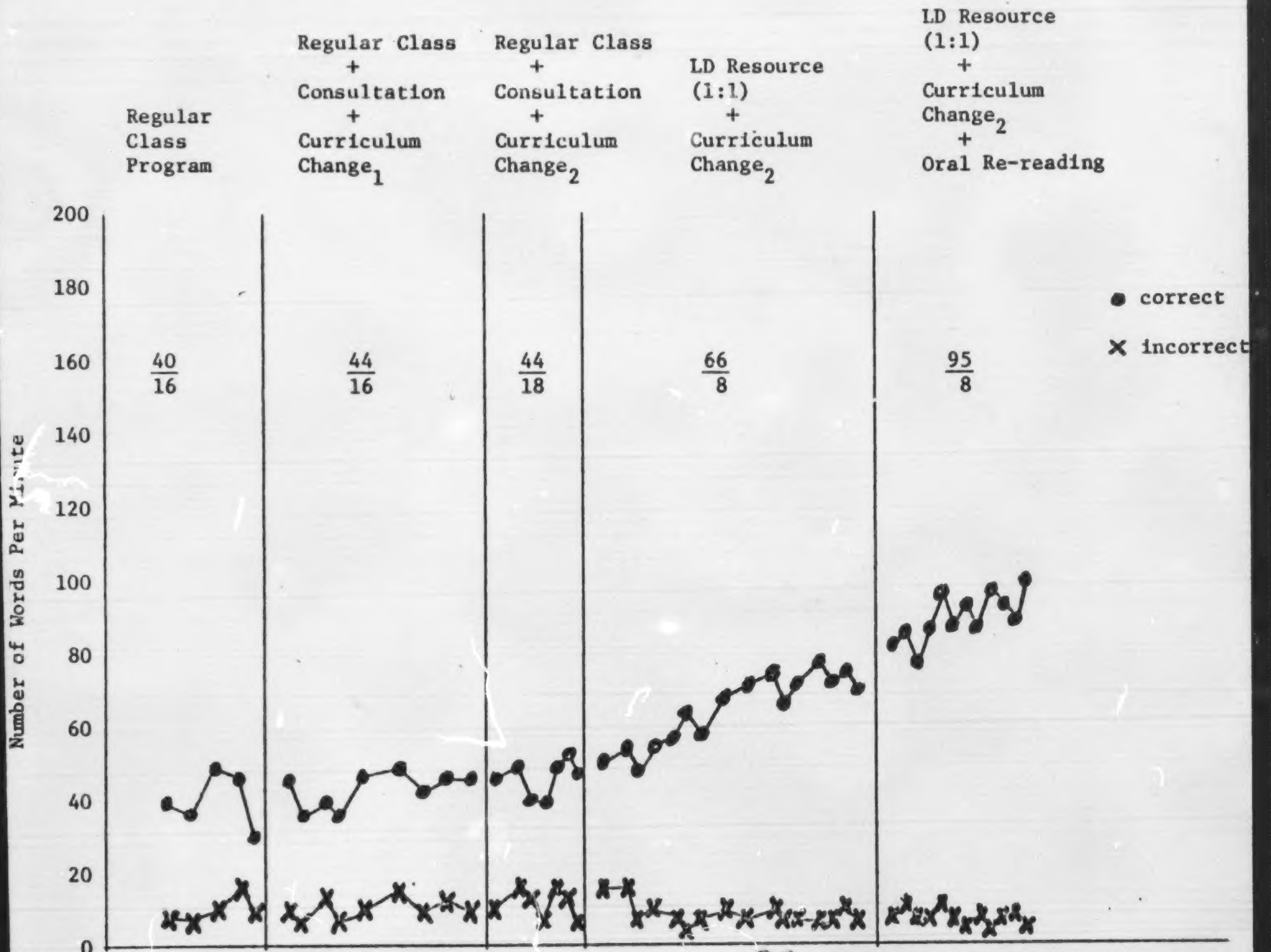


Figure 4

Reading Common Words in Isolation

Behavioral Research Methodology: A Reaction

Thomas Lovitt
University of Washington

I will organize my comments around the three features of the research program on which I was asked to make comments, and then talk about the six proposed research objectives. I will also present some ideas on applied research and where it seems to be going.

I will first comment on the knowledge base. It seems to me that there are at least three literatures related to this area of research. I think one would be the literature on precision teaching as set apart from applied behavior analysis. Although individuals in this area tend to be a little brash and careless, and often are poor writers, they are a rather creative bunch, and they move rather quickly from one idea to another. One has to be a bit careful in reading some of this literature. A second body of relevant literature comes from the applied behavior analysis camp - from the Baers, the Bijous, the Wolfs, the Bernbauers. A fine set of methodologies can be gained from them. I think sometimes they are a bit impractical in some things (such as school systems, political issues), but these researchers tend to be more applied than some people. Certainly, a third body of relevant literature that might be considered more strongly by Deno's group would be that in general education - the International Reading Association, the John Guthries, the Jay Samuels, and so on. They have been working on reading, mathematics, science, Minnesota state history, and so on, for a long time. I think that if Deno's group can pull together the literatures from those three areas - precision teaching, applied behavior analysis, and general curriculum, we will benefit greatly.

The second aspect of this research on which I will comment involves the rationale for the research. I couldn't agree more heartily with what Deno's group is doing. They are working on basic skills - reading, writing, and arithmetic; they are working in applied settings - schools, and they are dealing with a population that is very popular - learning disabilities. What could be finer?

The third point I will discuss relates to the potential contribution of the proposed research. My resounding approval here is based on the same reasons I gave earlier. Deno and his cohorts are studying fundamental problems and they seem to be doing it in a practical way in applied settings. I think some of the methods to which Deno referred (time series analysis and thinking about slopes over time and trends of data) can be implemented very readily in school systems if not all product situations.

I would now like to run through the six objectives of this research and make a few comments on them. The first objective is to measure behaviors in reading, spelling, written discourse, and social adjustment so that the effectiveness of LD intervention can be monitored. I hope that this does not mean that you want to find dependent variables that are highly sensitive to some of the old stand-by LD interventions such as the Frostig, walking rails, amphetamines, Distar, individualized cubicles, and DLM materials, which are commonly produced and ballyhooed. I think that might be a backward approach, and I really don't think that is what Deno had in mind at all. I think they are trying to identify important behaviors that we should deal with when we are trying to educate learning disabled children. In identifying important behaviors, you could just simply teach LD children the same behaviors that are taught to normal children, as has been done in the past. Or, you might identify the behaviors of successful adults and try to teach those to learning disabled children. Still another way would be to use a validation corps, an independent group of judges. Such a group would verify the existence of some deficit or maladaptive behaviors, then, after someone taught this set of behaviors, would re-evaluate the children and hopefully say, "Yes, now they are OK, they are non-LD." Conceivably then, you have dealt with important behaviors. I think it is time to really grapple with bigger designs, more important designs, and perhaps go beyond some of the individual charts that have been used for so many years.

The second objective of the research is to measure those behaviors (reading, spelling, written discourse, social adjustment). Deno referred to the controversy about measuring in terms of frequency, duration, percentages, and whether data should be collected on a daily, weekly, or monthly basis, and so on. There is also a great deal of controversy on how the behavior should be graphed - on the Cartesian coordinates that Stanley used or on semi logarithmic paper that Ogden would use, or on some other form of paper. There are great debates on these issues. There are several ways to resolve these methodological problems, but again, I think one way would be to have a validation corps, a group of judges. Of course, one consideration would be cost, such as the length of time required to teach one measurement system as opposed to another.

The third research objective is to determine who should monitor and measure the effectiveness of an intervention. Reliability is extremely important here. Often we use a friend, spouse, or roommate to be our second recorder of the data, if in fact we use a second person. We should be using highly trained, non-involved, and unobtrusive individuals to cross-check our measurement. In practice, I think that the effectiveness of intervention should be measured by an immediately involved person, such as the parent, the teacher, or even the child.

The fourth research objective is to insure that the data obtained in monitoring an intervention are utilized in improving an intervention. That's tough, but the answer is important. If data are obtained, they should be used in some way. They could be used to communicate, they could be used to validate, but they could also be used to make decisions. How to do this presents a real problem. Some studies have indicated that requiring teachers to collect and look at their data more often does not help them make more intelligent decisions. Often, it seems that specifying decision rules on the data chart is an effective way to encourage decision making. Dale Gentry, at the University of Washington, did an interesting dissertation along that line. He found that when they talked to teachers and prompted them to make decisions based on data, and gave social reinforcement for doing it, the rate of making decisions increased.

The fifth research objective is to identify systematic formative evaluation procedures that are logistically feasible within the context of school programs. The first problem is that no one has really identified the various systematic formative evaluation procedures, although Deno did to some extent. Second, I don't think that anyone has determined the extent to which formative evaluations are now used in school programs. Third, I don't know whether anyone has tested how often and with how many students these systems could be used in the school setting. If a teacher were to maintain four or five charts per day for 30 to 35 students, that would be 125 or 150 charts per day. It would be interesting to just test for limits on given sets of circumstances. However, one must also consider the politics and the realities of it all, which means we have to deal with busing, unions, collective bargaining, tenure, and so on, when developing a data system of any kind.

The last research objective is to determine whether systematic formative evaluation significantly improves intervention effectiveness.

One way to do this is to assess whether the system is more effective than nothing, whether it is more effective than another data system, and so on. You seem to plan to do this.

I think that the planned research is beginning in a different way. It's not a better way, it's a different way. I think that research problems are beginning to emanate from the schools. Previously, we might have sat in a library or in a cubicle and dreamed up a research idea. Then, after doing that research and coming up with significance at the .001 level, we would wonder why no one implemented it. Now, research is beginning to respond to the needs of people in the schools. I see this as one of the changes that we are going through in applied behavior analysis. I think that another change is in the measures that we take. I think we are seeing more data being gathered. More and more, I think that as researchers, we have to get the community, other people, involved in what we do. We are talking now about social significance as well as statistical significance. Finally, I think that the ways in which we disseminate and validate our programs are beginning to change. The important criterion in determining the value of research is the number of requests for information from teachers and other constituents; how many are used rather than just put on shelves? Were there any effects, and if so, of what magnitude? And, how many suggestions for research does the researcher get from teachers or other constituents?

Behavioral Research Methodology: A Reaction

Phyllis Newcomer
Beaver College

I would like to give my impressions of how I think this research focus can help the field of learning disabilities. I also want to talk about some of my concerns.

I find the behavioral methodology exceedingly attractive. I think it places the responsibility for assessment, for instructional purposes, precisely where it belongs, and that's primarily in the hands of the teacher. It gives the teacher a methodology for sequencing the learning activities in which the child needs to engage. The teacher is also disimbued of the notion that if a child has a learning problem, all she or he has to do is scream for someone to help and that someone will solve the problem. Therefore, I see many advantages to the behavioral approach, and I'm gratified that there will be, and there has been in the work of others, some very encouraging evidence that applied behavioral analysis is effective.

Now let me turn the coin around and tell you about some of my concerns. I am not associated with one specific approach to the diagnosis and remediation of learning disabilities. However, I have had experience with many children who have been called learning disabled, and I have developed certain impressions. You have indicated previously that part of your behavioral assessment approach will be for the purpose of identification. I was gratified when you said that essentially that purpose was not the focus of your efforts. If you look at any of the definitions that are currently prevalent, it is very difficult to fit your behavioral model for assessment, and particularly your basic assumption that learning disabilities is a natural product of the interaction between the child and his or her environment, into those definitions since they preclude calling children "learning disabled" when they are primarily environmentally or culturally disadvantaged. I think you really had to deal with that problem in order to justify your approach to identification.

I am far more concerned about the specific plans for demonstrating that this research into behavioral methodology is really going to benefit the group of children whom we call learning disabled. My interest is not only in learning more about behavioral methodology, but in learning more about learning disabilities. I have some qualms about your subject populations - children who have been labeled according to inconsistent criteria as LD. I wonder what conclusions you will be able to draw relevant to learning disabled children after you have completed your study. I can appreciate that you will be able to draw many conclusions relevant to children who have experienced academic failure. I recognize the importance of those conclusions. I realize that behaviorists prefer to avoid the issue of etiology because they regard it as insignificant. What is important is developing systematic programming to overcome learning deficits. And I can appreciate that position. When we try to operationalize the definitions of learning disabilities, we often want to tear our hair out, particularly if we are in an environment where we end up with 50 percent of our students as learning disabled. I agree with Tom Lovitt that the people who generate these definitions very rarely use an empirical base for them. However, while I appreciate your position, I am concerned, especially because the Institute is funded specifically for research pertaining to learning disabilities. We can get entranced with the behavioral approach and still end up with little pertaining directly to learning disabilities.

A related issue that I'm going to comment on is the person-centered approach for remediation as opposed to the situation-centered approach for remediation in a behavior model. I am firmly in favor of working with the youngster in the natural environment, the classroom. I have felt this way for a long time, even when the typical approach to a child's learning problem was to examine and remediate in isolation. I believe that we often got data that wasn't always as pertinent to the problem situation as we could have obtained by looking at components within the classroom. Although I firmly believe in the interaction aspects of any problem in any school situation, I would hate to see us reject the notion that a child may have a problem that is not induced by classroom related variables. The effect would be to shift emphasis and to maintain that there is no such thing as learning disability unless it occurs in a classroom situation. It's akin to what I've seen happen in the area of emotional disturbance. When I ask students in my introductory graduate course in exceptionality, "What do you think causes emotional disturbance?" they say that it is a socially determined variable. People are construed as disturbed because society sets standards and they deviate from them. In other words, they give a sociological perspective. I then ask them whether emotional disturbance is always relative to the social situation. When they say "Yes," I ask about schizophrenia. Eventually, they realize that not all conditions are relative to social situations. I would hate to see the pendulum swing in the field of learning disabilities to imply that everything had to do with the environment, and nothing with the child. I don't think that's true.

I also wanted to note that I'm very pleased that the focus of this investigation will be on basic skills. As Tom Lovitt said, what could be better than studying reading, writing, and so on. However, I noted that you propose to do some of the work with preschoolers. I am particularly interested in that population: I share Dr. Reynolds' concern that we cannot always be dealing with children after they become full-blown casualties in third or fourth grade. I am really interested in the kinds of data that would tell us something about these youngsters before they start failing and develop problems with basic skills. I wonder how extensive the exploration will be at the preschool level, and whether or not it will involve any investigations into oral language.

Finally, I want to make one more point that reflects my interactions with teachers. It is a recommendation really. You are proposing a very elaborate methodology. I know that you appreciate the importance of involving the teachers themselves in planning what is feasible in terms of the implementation of this methodology. I frequently have been in situations where these approaches have been implemented and teachers have cooperated. I think for a slight period of time, the new idea period, they enjoyed it and everything was fine. But then, when the researcher left, the new procedures were dropped, essentially because the teachers did not accept them as part of their role. One of the strengths of this study could be that the procedures established continue to be implemented when the research is ended.

Behavioral Research Methodology: A Reaction

Joseph Jenkins
University of Washington

Data-based program modification, the applied behavior analysis perspective, is probably effective by definition. It is a self-correcting approach which, when done properly, requires teachers to change instructional variables until the behavior being measured is affected. In such a system, less likelihood of short and long duration errors exists since teachers cannot persist with an ineffective intervention. That, it seems to me, is the essence of the behavioral approach on which you intend to capitalize in your research. With that preliminary comment, let me address myself primarily to your first research objective, which is to identify simple "thermometers" for measuring proficiency in reading, writing, and spelling.

While the objective may be quite clear, the issues that it raises are complex. The first issue is whether the measures really will help very much. Isn't it possible, for example, that if we made routine changes without measurement, we would obtain the same results? A second issue concerns the directness of the measures you might develop. Directness in this context refers to whether or not a test measures precisely what is being taught. Suppose other behaviors are changing, behaviors that may be more important than what your procedures measure? Conversely, if you take direct measures of the skills you are teaching, and make instructional changes based on the child's progress through your skill sequence, you are assuming that the skill sequence (or curriculum) is evenly and regularly progressing in difficulty. Only when curriculum segments are of equal difficulty can you conclude that your instructional procedures are responsible for variation in progress rather than the differential difficulty of skills within the curriculum. The issue, of course, is whether measures yield ordinal or interval data. If you develop measures based on progress through the curricula, the equal interval - equal difficulty assumption probably cannot be made.

Another problem related to developing measures is, it seems to me, that teaching-curriculum packages such as Distar, defy the kinds of measures you are proposing. You cannot easily make performance charts which represent proficiency on single tasks because the behaviors taught within these types of curriculum packages change from day to day. One option is to get away from direct measurement and go to some sort of indirect measure which can be given more regularly than standardized achievement tests. That approach requires choosing an ecologically valid, external measure, such as reading the newspaper. The advantage of an ecologically valid external measure is that it can be used to evaluate a change in curriculum. The only way to evaluate a curriculum change is with an external measure. Yet, there are problems when such external measures and a given curriculum do not match well. Your newspaper measure may not reveal growth, and you do not really know whether you are evaluating the entire reading program or the particular reading curriculum because the two are confounded. Thus, there are problems with both the direct measures, and the indirect or external measures, which you must address within your research. I have a feeling that we will probably have to go to some combination of both direct and external measures to provide the data-base for all the decisions we must make.

I hope that the external measures developed in your research are ecologically valid. Unfortunately, finding ecologically valid measures is not simple. The relationship between internal and external measures is important, however, because even though you might find an effect using a simple direct measure, it could be the case that nothing ecologically important has changed - that no effect has occurred on the external criterion measure.

Let me give you some examples of these issues as they exist in current research. In one study we were looking at two dependent variables - an isolated word measure and reading in context. Our intervention procedures affected the isolated word measure - a simple, daily measure - but, the intervention had no effect on the accuracy or rate of reading in context, which is a more ecologically valid measure. In another study, we set up conditions that affected oral rate and accuracy of reading. But, the increases in rate and accuracy did not affect reading comprehension. In a third study, we found that we could increase knowledge of word meanings, which is highly correlated with reading comprehension, but this did not affect comprehension. Such research suggests that even if you were to obtain high correlations in your validity studies, these could be misleading. You might identify good predictor variables (e.g., isolated word recognition), but changes in performance on such measures may not produce changes on your criterion variables.

As you develop your measures it would also be nice to find those that generate data with characteristically steep slopes, so that changes are quickly evident. Hively, for example, showed that growth on a cloze task typically changes so slowly that it takes too long to find out that a child is not making good progress.

I have one final set of comments related to the social behavior measures that you intend to develop. Steve Asher suggests that sociometric data are predictive of adult social success. For example, children with poor peer relationships show up more often on mental health rosters as adults, and they also have more run-ins with the law. Thus, sociometric instruments seem to have longitudinal predictive validity. However, sociometric instruments are hard to use. Asher recommends that one not use sociometric instruments more often than two or three times a year since they may be reactive. Thus, sociometric measures do not meet the qualification for simple repeated measurement that you seek. As an alternative, behavior analysts have tended to take rate measures on social interaction; these measures are easy to obtain. The problem is, however, that according to Asher, sociometric data and rate of interaction data do not correlate. Furthermore, there are no longitudinal studies on rate of interaction data to suggest that they predict success in adult life. So, data that we might be tempted to obtain regularly for assessing program effects on social development seem to be less than adequate. Perhaps the observational measures need to be modified somewhat to look at the rates of particular kinds of interaction.

Follow-up Comments to Reactants

Stanley Deno
University of Minnesota

I am going to make only some very brief comments. First of all, the issue regarding who we will work with and whether or not we will have anything to say about children with learning disabilities, I think, will rest on how we select our samples, on how we select the children with whom we will work. I suspect that our problem in identifying LD subjects will be no different from the problems of others as they try to develop their samples for doing research work. We will try to work particularly with children who others would identify as a group of children with learning disabilities. However, my basic assumption is that learning disabilities is a heterogeneous rather than a homogeneous concept. I hope that our findings will be applicable to any individuals we want to teach basic skills, so our research will be useful even if it turns out that there is no such thing as learning disabilities. Nevertheless, in our selection of research groups, I think we ought to try as carefully as possible to work with those children whom we currently refer to as learning disabled.

I appreciate Tom's comments about the validation core, about working with the teachers and the people who are the consumers and the users. I also appreciate Phyllis's comments because Phyllis is kind of a knee jerk teacher trainer or worker-with-teacher person. We have worked with teachers and I want to make sure that we continually do, so that our work will be useful and valid. I have great interest in Joe's comments, and would like to discuss many points further. Some of the problems with respect to covariation in our behaviors and criterion variables are of particular interest to us. We want to find not only high correlations, and good predictors, but also, corresponding changes in other important criterion variables.

Highlights of the Roundtable Discussion Following Deno Presentation

The Roundtable Discussion began with comments on possible political issues related to incorporating a teacher data collection procedure. The procedure is one which often is considered to be cumbersome. It requires time and effort on the part of the teacher. But, if the measures do lead to greater improvements in children's performances, it is relevant to ask whether teachers can afford not to measure. The political ramifications may be that such a procedure will require smaller pupil-teacher ratios, for example, which may not be realistic given the current state of affairs. It was suggested that the research should not be influenced by such "political" concerns. The first step is to identify and demonstrate the effectiveness of the data collection procedures.

Several individuals expressed concern over the difficulty of demonstrating transfer and generalization of skills that are measured and changed. It was suggested that the focus on skills does not come to grips with the pervasiveness of the educational problems. For example, even though skills may be developed within an individual, that individual still may not be considered to be a functional human being. You can change almost any specific behavior and teach almost any kind of competence if you focus a lot of high powered talent on it. But, what comes next?

It was generally agreed that the procedure of selecting a target behavior and applying massive intervention strategies to teach that target behavior is a fail-safe system. It doesn't fail because it is not allowed to fail. However, the research is concerned with contributing to the development of a fully functioning individual. The first step is to identify indices that are ecologically valid. Curriculum skills were chosen for study first because they relate to "cultural imperatives."

The research will attempt to identify behaviors that, when changed, reflect a relevant change in human growth and development. It was further suggested that an important aspect of the research should be to look at changes that occur in other criteria, rather than just in the behaviors under study.

Discussion then turned to the function of applied behavior analysis as a research tool versus applied behavior analysis as a teacher's tool. Behavior analysis appears to be a powerful research tool, for discovering instructional principles that are effective. But, is it a useful tool for the teacher? It appears to be cumbersome to use in the classroom. Teachers want things like more personnel and instructional materials, but they do not seem to express any need for more data to make day-to-day decisions on how to change instructions. What evidence is there to indicate that we need to use daily measures in classrooms?

One type of evidence mentioned was from research that contrasted teachers taking data with teachers not taking data. The research showed the augmented effects produced by data taking. It was suggested that such evidence is not sufficient. A more appropriate procedure would be to compare the effects for teachers collecting data versus those getting something else, such as more instructional materials.

Another suggestion was that the law now requires educators to show whether instruction has been effective for a child. There is an accountability basis for collecting data in the classroom. It was argued that such a function of monitoring achievement to document success is different from monitoring achievement to design instruction. The feeling was then expressed that while the law requires only monitoring to document success, researchers should take advantage of the requirement to determine whether monitoring can also be used to design instruction. Perhaps by requiring data collection, we can preclude some of the instructional failures that are so common in schools. It might be hypothesized that teachers will be more effective if they are more precise in the observations they make and the data they collect; whether they think they need more measurement or not.

It was suggested that an interesting line of investigation might be to look at the effect of data collection for teachers who, on their own, do not seem to have good "internal thermometers" of what is going on in terms of a particular child's progress. Also, it would be interesting to compare, at some time in the future, the effectiveness of a systematic data collection procedure against an informal evaluation procedure that is used in a successful instructional program, such as that used in the DISTAR reading program. It is not clear that systematic data collection will fare so well when pitted against a powerful instructional program. Such a study would have two groups using the instructional program, with one group using the formative data collection in addition to the instructional program. It was noted that the present research does have one study that represents a first step in that direction: given teachers presenting the same subject matter and using the same interventions, a no-measurement group, a measurement group, and a measurement decision group will be compared.

The point was made that the Institute is focusing on the methods appropriate for children with learning disabilities, not on the investigation of applied behavioral analysis. The orientation of the research to success in education was commended. It was noted that the researchers are trying to define whether a measuring system is useful in terms of the way children learn.

Highlights of the Open Discussion
Following Demo Presentation

Discussion began with comments on the importance of conducting research before attempting social validation of the technology. It was noted that oftentimes technology precedes consumer demand. For example, miniaturized transistor circuits were available long before there was a big demand for desk calculators. The proposed research is designed to determine whether an available technology works for a given population. Social validation requires that we have some answers. While some aspects of behavioral principles have been validated, the use of this measurement technology for increasing teacher decision making has not been validated. The first charge should be to experimentally validate that function, before an attempt is made to socially validate it. Concern was expressed that teachers should still be involved in the experimental validation process.

The question was posed of who will determine what target behaviors to work on. Is successful learning merely whatever the teacher identifies? For example, many teachers work on b-1 blends even though the child never misses them when reading in context. There is the potential problem of measuring success in an on-going program when that program is completely inappropriate given the complexity of the youngster's learning problem.

It was agreed that too often the tendency has been to be curriculum bound with assessment procedures. The goal of the present research, however, is to develop ways of taking regular data that will be curriculum free and independent of the specific skills being taught.

Although it was suggested that issues related to teaching irrelevant behaviors or simple behaviors rather than complex behaviors are relevant not only to the behavioral approach, others suggested that the problem of measuring simple versus complex skills is particularly important. Teaching oral reading, which can be easily measured, does not improve comprehension. Thus, reading comprehension must be measured to see whether changes are occurring. But, it is difficult to measure a child's reading comprehension in less than half an hour; doing such a measurement every day involves time away from reading lessons and other instructional activities. When you want to teach complicated things and you also want to measure complicated things, you have a trade off between testing time and teaching time. The question then becomes: how much teaching time are you willing to pay for measurement time? The question of measuring and changing behaviors on simple dependent variables versus complex dependent variables immediately gets into the problem of time. The measurement system is not necessarily justified on the basis of its demonstration with simple independent variables. The approach of the present research, to look at new time frames for measurement, new sampling systems for measurement, a more open system about how to measure and when, is very important in the light of the time problems that do exist.

In relation to this point, others noted that although we currently do not have quick ways to measure more complex skills, the possibility of finding such measures does exist. It was noted that the research project seems to have two major thrusts - first, to see whether the measurement approach works, and second, to see whether the approach can be applied in the area of learning disabilities.

Another individual noted that the proposed measurement system is precise in the sense that an observer records the frequency of a precisely defined behavior. The measurement also occurs at the time the behavior occurs. The problem with the more global judgments typically made by teachers is that they are made after the behaviors have occurred, and there are distortions of judgment as a function of memory. In order to compare teacher judgments with more precise measurements, to find out whether the precision is necessary, the research should equate for the time at which the judgment is made.

Another member of the audience suggested that the focus on teachers recording data may be too narrow. The teacher does not have to be the only person recording behavior. Children also can record behaviors. The concern with teacher time for measurement may simply reflect the fact that we are not yet being creative enough in identifying the many different possibilities that exist for recording data.

The importance of methodological concerns was proposed. It was suggested that control conditions and control groups should be handled separately. In looking at control conditions, the research should look at what happens with the use of the measurement system versus what happens otherwise. Also, the research should look at the amount of time a teacher devotes to intervention. Thus, if a teacher in a control group, not using the system, has to spend an equal amount of time taking notes, for example, then there are control conditions in the control group. The research might also benefit from employing a within-teacher design for control. This would involve having the teacher introduce the measurement system following a period in which the system was not used. This would also provide a form of social validation, in which the teachers could say whether the addition of the system was helpful and whether they would continue to use it if the choice was left to them.

Another member of the audience agreed with the point that it might be possible that the significant effects of daily measurement are related to the fact that the measurement system gets the teacher to attend constantly to the program. There are probably many non-specific things that can be done to focus the teacher's attention on the students. However, daily measures do have the advantage of being microscopic and thus perhaps not so subject to halo effects and memory problems. Yet, to get the teacher involved in a way that constantly points the teacher to the task may be quite a different sort of effect that has nothing to do with the problem of daily measures.

It was then noted that perhaps the best single variable by which the teacher's behavior should be controlled is student performance. While other methods may increase the teacher's attention, it seems reasonable to try to develop a measurement system that gets the teacher's instructional behavior under the control of the student's behavior, so that as the student is improving or not improving, the teacher will make changes.

The discussion concluded with comments on the relation between teaching and assessment. It was noted that the research project views teaching as an assessment technology as well as an instructional technology. It is important to feed data obtained through instruction into the decision-making process. Teachers probably are the people who ought to make those assessments for the most part, and feed information from those assessments back into the decision-making system.

Summary and Commentary

John T. Guthrie
International Reading Association

The papers, discussions, and reactions in this conference have reviewed a great deal of literature on assessment of children with learning disabilities. A basic purpose of this conference has been to state the important problems in the field that require research and propose investigations of these problem areas.

My commentary is centered around two questions: (1) Are the problems that have been stated well-formed? and (2) Will the solutions to these problems benefit the education of children with learning disabilities? It is interesting that these two questions have received more and more attention lately in the research establishment. Gaining consensus in the research community about what the important problems are has been more common in the National Institute of Education in the last few years. For example, ten task forces were formed to develop the important research issues in reading, the outcome of which has been the Center for the Study of Reading at the University of Illinois that is focusing on comprehension of children at the intermediate grade level. The second question, about educational benefit, has received more attention as federally funded research has become increasingly mission oriented. Regardless of whether this orientation is healthy, it raises the stakes for studies that seem to promise immediate improvements in education.

Framework for Assessment

I have pieced together a somewhat shaky organizational framework to help me understand the relationships between the two main papers, sections within the papers, and the reactions of the critics. It happens that this framework also points to missing data that are crucial from at least one viewpoint, though they are superfluous from at least one other viewpoint.

The framework consists of four components of the process of attempting to improve education for learning disabled children. These components are: (1) referral; (2) assessment; (3) diagnosis; and (4) teaching. Referral pertains to the event in which some person, usually an adult such as a teacher, parent, or principal, suggests that a certain child in a classroom or school needs help. Needing help usually refers to a child's failure to learn a basic skill in a manner that makes the child unmanageable by the teacher or the child's behavior that disrupts learning for the children. Assessment refers to the administration of tests that are intended to determine the child's achievement, aptitude, or social and attention behaviors in comparison to other children or to the demands of the curriculum in which the child has been learning or to the requirements of the social context in which education is taking place. Diagnosis refers to statements about the child that delineate the problem or to a classification of the child into a category based on a familiar clustering of scores on the assessment tests. And, teaching refers to different educational activities, management systems, or interventions by specialized personnel, such as tutors or counselors, that are intended to solve the problem as described in the diagnosis. It is not hard to see that decision making may occur within each component or at the linkage points between the components.

Validity of Components

To determine the adequacy of a component, or the decision-making processes entering into the component, we need a criterion for each one.

Essentially, we need to know how to validate each component. One reasonable criterion that can be used in the validation of these components is that each stage is valid if it leads to valid decisions in the next stage. For example, a diagnosis is valid if it leads to a teaching program that improves the child's achievement in basic skills by some appropriate standard. In turn, an assessment procedure is valid if it provides sufficient information to enable a reliable diagnosis to be made. In turn, a referral is valid if it leads to new assessment information that is important for the child's education or the confirmation of a suspicion regarding a child's learning problem.

It can easily be seen that this system hinges on the validation of teaching. The dependent variable for decision-making research is the validity of teaching. For our purposes, teaching may be considered valid when a child learns rapidly and behaves appropriately according to some agreed upon standards in the educational program in which the child is located. Within this system valid teaching assumes, at least at some gross level, an aptitude treatment interaction. That is, there must be a treatment or a teaching system that is effective for a defined subpopulation #1, but less effective for a subpopulation #2. At the same time, a different treatment or teaching system is effective for subpopulation #2, but less effective for subpopulation #1. If this is not the case and teaching cannot be validated in this sense, then the validity of the other components of the decision chain cannot be determined.

The authors of major papers and the discussants in this conference have taken different positions about the relationships among these components. For example, Keogh states that the real issue under consideration in this program of research is decision making within the context of educational programming. In other words, to use the language of the framework, she is emphasizing the validity of the components and the processes that enter into them. Specifically, she emphasizes the linkage between assessment data and educational decisions about LD children.

Keogh also shares her belief that the assessment and identification of learning disabled children is an important goal, deserving of long-term and systematic investigation. Do we know that this is true? How do we know it is not a waste of money to assess and identify these children? How do we know they do not perform more effectively in regular classrooms? Do we have evidence that interventions are effective?

Not long ago, I was informally observing a fifth-grade classroom and studying three or four of the children who did not seem to read well enough to keep up with their lessons. The teacher, Mrs. D., pointed out one child to me who she thought was bright and conscientious but was one of the lowest readers in the class. Partly by way of description, and partly as a request for advice, she said, "I don't send him to Mr. X [the reading specialist for the school] because I think he learns more in my classroom." What is a scientifically based reply to this teacher, I asked myself as the blood drained out of my face? In a moment this teacher has jettisoned the referral-assessment-diagnosis-treatment process for children with learning disabilities because she didn't see any validity to the teaching component. Until we can tell her the probability that this child's reading will be improved by an intervention, and tell her the amount of that improvement, there is little rationale for the practice (or the investigation) of other components in the decision-making process. With substantial respect for her judgment and little scientific evidence to the contrary, I told Mrs. D., "You're probably right--he does learn more in your classroom."

In the first area of research discussed by Ysseldyke, the adequacy of norm-referenced data for prediction of success in a highly systematic instructional program is being investigated. Following this study a weighted combination of certain tests will be identified that can predict those who learn a great deal from those who learn very little from this teaching program. What can be done with that formula after it has been devised? One seductive implication that we should avoid is that children with a profile similar to that of successful learners in that teaching program should be placed within it. It may be that the same profile predicts successful learning in other programs as well. A control study is needed in which the same test data are available for children who learn a great deal or learn very little in a less systematic teaching method. However, if there were differences between the formula for the control group in flexible teaching and the formula for the experimental group in highly systematic teaching, the useful result will have been turned up. Teaching will have been validated in one sense, and a rare aptitude-treatment interaction will have been discovered.

Reliability of Components

If a decision-making chain, such as the one I have outlined here, is to function effectively, each of the components must be reliable. By reliability, I am referring to consistency in performing its function in the decision-making chain. How reliable is the teaching component? In other words, do we know how consistently an educational program of a certain description, say the Distar system taught by an experienced teacher, will meet some agreed upon standard for effectiveness, such as one year of gain on a standardized comprehension test in ten months of teaching for a given population of children? This reliability could be stated in the form of the probability that any child from age seven to fifteen with normal performance on an intelligence test such as the WISC and two or more years' deficiency in standardized reading test performance will make such a gain. To my knowledge, there are few data to this effect, and although collecting such data has hazards, they would provide crucial information to the study of the decision-making process.

It has probably occurred to you that teaching LD children cannot be validated without being able to locate LD children in some consistent way. Reliability of decisions in the diagnostic component seems particularly important, and studies of it have been proposed by Ysseldyke and Lovitt during this conference. In Research Area 3, one study is proposed by Ysseldyke to examine the differential diagnosis of LD and non-LD children. That is, 60 professionals will be asked to sort 120 cases of individual children into those who are LD and those who are not LD, based on 22 pieces of test information. The percentage of cases on which, say, more than 90 percent of the clinicians agree (about this binary decision of LD vs. non-LD) will give some evidence about reliability of diagnosis into these broad categories.

It seems to me that reliability of diagnosis is a central problem, since most educational placement decisions are based on it. LD children often receive tutoring, milk and cookies, and direct, daily measurement; whereas, non-LD children seldom receive any of these. The essential questions on this problem are: (1) Do clinicians agree with each other on the classification of children into LD and non-LD categories? Designs such as those presented at this conference may work to answer this question. (2) Are teams of professionals reliable; that is, do they agree with other teams on the classification of children? (3) Are clinicians or teams of clinicians consistent over time? That is, do they give the same diagnosis on the same or highly similar cases on two different occasions? I might add that I would predict that the reliability of diagnosis will be inversely proportional to its usefulness for teaching. That is, reliability of LD and non-LD judgments may be reasonably high, but their value for teaching is limited. To know that a child has "a reading problem" does not provide a tutor a large step forward, and the information could nearly always be provided by the referral source--the teacher or parent--in any event. On the other hand, information that the child needs to be taught how to divide words into syllables for decoding, to expand his or her knowledge of word meanings, particularly in the area of abstract nouns and verbs, and to expand his or her perception of the thematic structure of stories by adding the detection of resolution to his or her present abilities in detecting beginnings and climaxes may be useful to a teacher. However, agreement among clinicians that these (or some other) are the appropriate teaching goals is not likely to be high at this point in the development of the field.

The other investigations presented by Ysseldyke pertain to the factors that enter into decisions in the other components in the decision-making framework I have sketched. For example, Research Area 4 entails sending questionnaires to decision makers, including principals, special education administrators, and child service demonstration centers. The respondents will be asked to indicate which of the components are included in their decision-making programs, the particular operations used, and the personnel involved. Research Area 5 involves studying placement team decisions through direct observation and tape recording. The second research area, involving interface with the computer, will study the effect of certain variables, such as sex, economic status, physical attractiveness, and academic vs. social difficulties on diagnostic decisions. These studies involve examining the variables that influence a diagnosis. It seems to me that a case could be made for deferring these studies until such time as the diagnosis can be validated. Only after we know whether these educational decisions make a difference for the teaching of LD children and which of these educational decisions make a difference do the decisions themselves merit inquiry. If a decision is pointless and ineffective, then the study of how it is made is equally pointless and ineffective.

Behavior Analysis Approach to Assessment

The major components of the decision-making process, consisting of referral, assessment, diagnosis, and teaching, are viewed very differently by the behavior analysis group, including Deno, Mirkin, and Shinn, Lovitt, Jenkins, and others, than they are by the psychoeducational studies group of Ysseldyke, Thurlow, Keogh, Salvia, and others. The behavior analysis school claims that the decision-making chain, that I outlined before, of referral from the teacher, assessment by examiners, diagnosis by professionals, and teaching by a different specialist is inoperative. This decision-making program is dismantled by Deno's rendition of Lovitt's view that the teacher must continuously evaluate different techniques which are used to solve individual problems, and no single instructional technique will ever be appropriate for solving all the problems presented by children who have difficulty academically. Teaching is primarily a matter of trial and error, a matter of continuous, and sometimes drastic, adjustments of lessons to cope with changes in the learning of the child or a group of children. Therefore, the behavior analysis group proposes to develop formative evaluation systems which will detect errors in the instructional problems quickly. In other words, in the terms I have been using, because the teaching component cannot be validated, the rest of the system does not need to be examined.

The behavior analysis people, on the other hand, also make decisions about teaching LD children, and their decisions appear to be like those of everyone else. The decisions include: (1) Referral--which consists of the decision to conduct a decision-making routine. This decision is often made by a teacher based on his or her own judgment, or based on Lovitt's judgment about daily measurement. (2) Next, the teacher assesses the child by giving a measurement of behavior, for example, a list of 20 words that the teacher would like the child to be able to read orally. (3) The teacher will diagnose the child, which is to say, decide whether the child has learned or not learned the list according to some previously set standard, such as one word per second correct and 1/10 of a word incorrect per second. (4) The instructor engages the teaching component. The instructor decides that if the child has not learned List A, then List A will be taught. Contrariwise, if a child has learned List A, the educational program will proceed to List B. These decisions, rather common in behavior analysis, include referral, assessment, diagnosis, and teaching in their own way. There are decisions within each component and between each component, and many options may be found. The processes are occurring at the level of a lesson in the behavior analysis scheme, though they are occurring at the level of the syndrome for the psychoeducational group.

Validity and Reliability

How are the issues of validity and reliability for the decision-making components addressed by behavior analysis? The validation of teaching is always the most difficult problem to confront, and despite its elusiveness, I'll attempt to treat it briefly. The criterion, mentioned previously, was whether an educational intervention could be shown to be uniquely useful for one subpopulation of children. A more appropriate criterion for the validation of teaching for behavior analysis regards decisions made about what to teach and when to teach it. For example, if a child performs below standard on Word List A, normally Word List A is taught until it is learned to mastery (in some sense of the word). If a child has not learned Word List A, the teacher seldom goes to List B for instruction or to Story X which contains many words from Word List A. What is needed in the behavioral analysis framework is an examination of the validity of this teaching decision. If the long-term goal is for children to learn Word Lists A, B, C, and D, and to read Stories X, Y, and Z, what is the most appropriate teaching strategy? It may be to learn List A to criterion then to read Story X (which contains words from List A), read list B to criterion, read Story Y, List C and D, and read Story Z. But there are other options--such as not learning word lists but only Stories X, Y, and Z, plus Stories A, B, and C, which also include the same words, or learning A, B, C, D, Word Lists and Stories X, Y, and Z to a low criterion, and then learning them all again to a high criterion of proficiency. The history of education is resplendent with eloquent armchair debate about these options. But behavioral analysis can develop data regarding them, as their advocates are well aware.

The question about validity of diagnosis essentially amounts to the issue of how much learning (of Word List A or Story X) is needed to judge a child as proficient. There is substantial evidence that mastery learning does not apply to acquisition of word recognition. Children learn orthographic rules, structure of words, pronunciation of unknown words, and other word recognition skills continuously from first to twelfth grade. Mastery of one word during one lesson at one age does not occur by any reasonably stringent criterion. Although the child may pronounce a word correctly within a given lesson, his or her speed of pronunciation will increase over the year, his or her knowledge about the varied meanings in varied contexts will increase, and his or her ability to linguistically predict its presence in a clause will improve with general language skills. Both substantial data and simple reflection allow that we have all increased our knowledge of meanings of words, such as war and peace, or our reading of simple documents, such as the yellow pages. And so we are required to evaluate the concept of mastery and sequence within the component of diagnosis for applied behavior analysis programs.

The validity of the assessment component in behavior analysis programs leads to the question that Deno posed as central: what should be measured to determine whether a teaching program is effective? Deno and his associates indicate that measures they have used or propose to use consist of high frequency words in isolation or oral reading of words in context. I confess to say that it was a relief to find out that reading words, which is a major problem for these children, would be assessed by giving children words to read.

The question I am raising is whether the task of oral reading is an appropriate measure in the education of learning disabled children. It is elegantly simple to ask a child to read aloud, count his errors, and time his performance. The record is sensitive to reinforcement contingencies. Oral reading seems to be a valid dependent variable. Since the child is reading aloud, it appears that the behavior is under direct observation. But is it? Presumably, the fundamental purpose for reading is not to make sounds but to derive meaning from print. In a series of studies, Lovitt and Eaton found that reading comprehension was utterly unrelated to dramatic changes in oral reading rate and accuracy that were wrought by clever reinforcement contingencies. The validity of the proposed assessment device needs to be examined. I very much like the concept of validation corps that Lovitt proposed. The top of his list for issues to be validated by the research community is the importance of the response. This relates to whether the behvaior being modified is significant. In the case of word recognition, the importance is moderate. Reading cannot occur without this skill, and yet this skill is by no means sufficient for a student to understand a social studies text or to enjoy a bubble gum wrapper.

Reliability

The issue of reliability of components of referral, assessment, diagnosis, and teaching in the behvaior analysis approach will not be discussed now, partly for lack of time and partly because behavior analysts are sensitive to this issue.

Closing

In closing, I would like to refer to an observation by William James (McDermott, 1967). He said,

> Our knowledge grows in spots. The spots may be large or small, but the knowledge never grows all over; some old knowledge always remains what it was....Our minds thus grow in spots; and like grease spots the spots spread....The novelty soaks in; it stains the ancient mass. (p. 418-419)

As I see it, the research proposed at this conference will provide a spot of new knowledge in the field of learning disabilities, and my hope is that it may visibly stain to a more favorable color the ancient mass of literature in this field.

Reference

McDermott, J. *The writings of William James*. New York: The Modern Library, 1967.

Summary and Commentary

John Salvia
The Pennsylvania State University

I have two major points that I would like to make. One is on the compatibility of the two teams' work, and the second is on the general practicality of the Institute. Then, I will present a series of picky comments, and finally, some more fundamental concerns.

I think that this is one institute that is divided into two teams and, to some extent, the two teams' work should be integrated and compatible. Jim mentioned four major areas of assessment: screening, classification, program planning, and program evaluation. I think that the two teams have divided the purposes of assessment between them. I think the division is fairly reasonable. Deno's team is concentrating on, in their words, "the effectiveness of the applied behavior analysis approach to assessment in programing for children with learning disabilities." I would call that program planning and program evaluation. Ysseldyke's team is concentrating on the problems of classification. This seems to me a very reasonable division of labor. But, as I tried to put the two together into an overall framework, I found that the teams ignored the work of each other in many areas, and I think in one spot, at least, the work is contradictory. I'd like to run through those. I think the solution, by the way, is fairly simple. The two teams need to sit down and talk a bit.

Both teams are interested in the utility of the data collected for placement and classification. Deno's team devotes considerable attention to this issue. For example, they address the issue of person-centered versus situation-centered approaches to assessment. Ysseldyke's team is so interested in assessment and classification that their fifth research area is

"Ecological Research on Team Decision Making." Although six studies are listed by Ysseldyke's team, the concerns of Deno's team are ignored. In the computer simulation research of Ysseldyke's team, they could, but do not, incorporate options for person-centered data versus situation-centered data. In perspective 12, Deno's team considers personal and social problems as they are related to the identification of pupils as learning disabled. Yet, I do not find any mention of personal-social variables in Ysseldyke's retrospective studies in Research Area One or their simulation studies in Research Area Two. The comparison of LD children with children who are not labeled LD but are failing in school (Ysseldyke's Research Area Three) uses the Piers-Harris Self Concept Scale and the Peterson-Quay, but no direct observations. In all fairness, it should be mentioned that the questionnaire studies in Research Area Four do ask about personal-social variables. I will not talk about Research Area Six now, other than to indicate that although observations of personal and social behavior could play an integral part, maybe a crucial role, in bias following assessment, this aspect is ignored. Similarly, Deno's team has not mentioned some of the important issues raised by Ysseldyke's team. For example, Ysseldyke devoted considerable attention to the notion of bias in assessment. Yet, I find the issue addressed by Deno's team only in connection with situation-centered assessment, and there I think it was dismissed fairly casually. About the only mention I found was in the statement:

> If during an interview with the teacher who has referred a student, the resource person detects some bias operating with respect to the teacher's identification of this particular child as a problem, an opportunity exists to address the teacher's biased expectations as well as or instead of the student's behavior.

In addition to the literature on biased standardized tests and biased ratings, to which Ysseldyke has attended, there is a small, but growing literature on biased behavioral observations. In fact one issue of JABA contains a paper by Al Kazdin which talks about problems in behavioral observation; simple counting is a biasable phenomenon. One particular problem that I have been concerned with is observer drift. There is a tendency for observers to change their definitions over time and not record the same behavior. In an institute in which a substantial portion of one team's work was directed toward bias, some concern with bias should be reflected in the other team's work. Another bias study with behavioral data is a study by Jack Neisworth. He found that labels of hyperactivity affected counts of out-of-seat, hyperactive kinds of behavior.

The last point I wish to make in connection with the compatibility of the two teams is that neither team proposes to study the effect of the label "learning disabled" on program planning, program evaluation, or program implementation for learning disabled children. I would think that would be a very big issue in terms of bias following assessment, and one that could easily be incorporated. The preceding points have dealt with integration, and I consider them to be merely omissions. Each team has failed to take into account the perspective or the data base of the other team. Such omissions could jeopardize the validity of various experiments.

I think that there may be a conceptual contradiction in the work proposed by the two teams. Deno's team is using a highly-structured, behaviorally oriented approach to instruction. They believe it will work. There are two quotes that are illustrative:

> The primary goal of the proposed research is to develop formative evaluation procedures, based on time series data analysis which, when used by teachers, improves the effectiveness of attempts to modify programs for children with learning disabilities.

and:

> The research activities proposed promise to provide the means for improving programs of instruction provided for students with learning disabilities.

On the other hand, Ysseldyke's team defines as learning disabled only those students who do not succeed in such a program. As they put it, a systematic instructional program

> based on task analysis of learner skill development and employing continuous daily measurement...has been highly successful with the majority of students, but has been unsuccessful with many others. We identify the former as instructional casualties and the latter as true learning disabled children.

Now it seems to me that if the true learning disabled child won't learn under the structured situation, then Stan has a problem. If, on the other hand, they do learn under a systematic program of instruction, then Jim has a problem. And, I think that if they don't get together on it, we have a problem. Basically the programs fit together; there are not a lot of problems or difficulties. But, I do wish that you two would address those issues.

I have some picky points that I just can't resist mentioning. I have four points for Jim and one for Stan, but the number is not indicative of the weight. Some of my concerns have been raised earlier, but I would like to drive them home. In Research Area Three, Ysseldyke talks about differential diagnosis of LD and non-LD children. They are going to compare children who have been diagnosed as learning disabled with children who have been diagnosed as non-learning disabled. They are going to give a whole battery of tests to those 120 children, 60 in each group. As best as I can tell, the data on LD subjects are going to be collected sometime after the kids have been in special education. My question is, how will you attribute differences - to special education treatment, or to differences at time of diagnosis? I think it is going to be a very difficult project to interpret. I'm reminded of some of the neurological studies that were done where people tried to find locations of tumors, waited twenty years and performed autopsies. If there were no tumors, and the diagnostician said there were no tumors, everything was just fine; if there was a tumor on autopsy, one was always open to question as to when that tumor occurred. Was that the same tumor that they found 20 years ago? Perhaps more fundamental is the question of the inter-rater agreement of diagnosis of learning disability. In a follow-up study diagnosticians are going to look at the records and differentiate children who have been identified as learning disabled and not learning disabled. My question is, what happens if they cannot make a reliable distinction when you go to compare learning disabled with non-learning disabled children? You have a dichotomous-criterion variable that by your own research will demonstrate no reliability. If you do not find group differences, do you attribute it to faulty classification procedures or to the measures? Similarly, if you do find group differences, but diagnosticians cannot agree that one group is learning disabled and the other is not, what do those differences mean?

In Research Area One, you will develop a definition of successful and unsuccessful students in order to compare kids who have profited in an LD program and those who have not. It is not as easy as it appears to make a cut in there, to decide what constitutes successful and unsuccessful.

A number of studies that I had something to do with were reviewed; the general conclusion was that these studies do not offer a lot of insight or validity. I think they do. I think that in your computer simulations, in your questionnaire research, and your observation research, you are doing much the same sorts of things that we did with the attractiveness work. We asked teachers what they thought, we went out to look at what they were doing in terms of process, and then we looked at the outcomes.

My last point for Jim is that I have some concern about the interactive effects in the studies that deal with questionnaires, computer simulation, and observation. As I understand it, they include the same psychologists. I wonder how much participating in one study will affect their behavior in the next study. I recognize that there are many advantages to studying the same people over time, but there are also some disadvantages.

The last picky point is for Stan. I'm really amazed that he is going to set off to develop a new set of achievement tests in all of those areas that are based on observational data. I think this will combine the problems of both behavioral observation and standardized testing. You are going to have to worry about both inter-rater reliability and regular kinds of reliability.

Practicality is a more general issue that I would like to discuss. The question is, how useful is this Institute going to be? That is, will this Institute produce research that will make a difference in the identification and treatments of learning disabled pupils? My answer is a wholehearted "maybe." If everything goes well, I guess it could. My pessimism lies primarily in the area of decision making. I think you will find a lack of main effects, a lot of error variance, inconsistent judgments; I do not think there will be a consistent pattern. My lack of confidence does not stem from doubt about the team's capacity to do good research or the methods that they will be using, it's just the state of the art. I think that we are intellectually bankrupt and we are trying to cope with a system that is not very scientific. If we are supposed to be a scientific discipline, then we should be able to agree upon what we are observing. In the case of learning disabled children, we should be able to agree that x, y, and z children are indeed learning disabled. I cannot imagine a group of chemists arguing over whether or not a sample contains iron.

I disagree with the notion that we should try to bury definitional issues. I disagree with the premises on which the classification of learning disability are based; I disagree with the definitions. This doesn't leave much. The first premise with which I have a great deal of difficulty is that mental processes exist as anything other than descriptions of behavior. Certainly, they do not exist as things than can be stretched and formed in the sense that most people are talking about, and in the sense that the definition is used. The second premise with which I have a great deal of difficulty is that in the joint occurrence of behaviors, process errors and academic errors, anybody can separate a cause from an effect on one child. The problem of attribution is key here. The question asked is, what is it that has caused this child's failure to achieve? I think we engage in that kind of exercise every time we diagnose a learning disabled child. Deno spoke to that point well in his paper. Speaking to it does not make it go away. Now, I would not argue that for some kids, attribution is quite possible. But, we seldom have clear-cut evidence. The third premise that I have a great deal of difficulty with is that discrepancies in an individual profile are either reliable or diagnostic. I see no reason to believe that scores that do not correlate perfectly should be the same. It's just that simple. But, most fundamentally, I reject the notion that non-educational variables can be used to classify educational problems or to provide insight into the delivery of educational programs. As Deno and Mirkin wrote in Data-Based Program Modification: A Manual, "Handicaps themselves permit us only to speculate on why a pupil is difficult to teach. Merely identifying a handicap often obscures the pupil's instructional needs" (p.55).

Later on the same page they say, "From our point of view, then, the problem is never a child's handicap, it is always the discrepancy between desired and actual performance and progress." There are many relevant pupil characteristics, non-educational variables, that cannot be ignored; we know about the effects of hunger and malnutrition, we know what illness does, and we know what uncorrected physical defects do. Yet, if we have a hungry child, if we have a child with a severely limited memory ability, our job is still to teach the child.

The second problem is with the definition itself. I have been trying to get a fix on who learning disabled children are, and I would like to share a folksy perspective with you. Learning disabled children are children who, for no apparent reason, are performing poorly in school. "No apparent reason" in that these children are free from debilitating intellectual handicaps, emotional problems, sensory problems, motor handicaps. The children are culturally similar to most children, but something has got to be causing their poor performance. As I see it, there are roughly three things than can cause their poor performance.

The first one is tests. Norm-referenced tests are designed to rank individuals in terms of relative levels of mastery, to discriminate among pupils, to separate pupils along various continua. Test questions that do not discriminate among individuals are not included in the test. Test questions for which low-scoring children answer better than high-scoring children are deleted. The job is to spread the children. Achievement tests and intelligence tests, as well as tests of particular processes that are used to identify learning disabled children are, indeed, norm-referenced tests. That leaves us with the conclusion that some children must flunk. If we define normal as 25th percentile and up, for either achievement or/and intelligence, and we say you have to have normal intelligence to be learning disabled and you have to have abnormal achievement to be learning disabled, and if we assume that the tests are independent (which they are not), then we will still come up with about 18 percent of the population as meeting that criterion of normal intelligence and poor achievement in each achievement area tested. The point is that the tests are designed to identify children who are behind, and if you make a definition based upon low performance in two areas, you are going to have children behind no matter what their internal state. So tests, I think, can cause us to have a LD population.

Second, poor teaching. We don't like to admit it, but there are poor teachers out there. Bob Smith and Jack Neisworth use the term "teaching disabilities." Children with normal ability may not be achieving simply because of incompetent instruction, repeated by various teachers.

The third explanation for why pupils are not doing well in school [illegible] no apparent reason is for a reason that we cannot see - something inside the child. There may be something that's wrong with the children, something inside. I think our social policy will not allow us to provide compensation for victims of poor teaching, or inappropriate testing, or who are poorly motivated, but it will provide money for children who have something the matter with them. If we are going after the children who have problems inside and aren't making it in school, then to operationalize that definition is fairly simple, at least conceptually. Somebody needs to teach the children under highly motivating conditions. If a child doesn't learn, doesn't meet criterion within a specified number of trials, we have a functional definition of learning disabilities, much the same way

Ysseldyke indicated. The profession has not opted for such an approach. Most definitions of learning disability are scientifically useless; they are empty jargon or prattle. The following is part of the Federal definitions: "A process deficit or dysfunction in one or more of the basic psychological processes involved in understanding or in using language." This is included so that the pupil is not a victim of poor testing or poor teaching. I would like to see an enumeration of the processes; if the child has a process deficit, let us list the processes that we're considering. Unfortunately, the enumeration of such processes is seldom given because they depend upon one's theoretical orientation, and the theoretical orientation of the person who has the power to define has been very important. But we could talk about self-actualization, need states, perception, perceptual-motor integration, conceptual tempo, attention, binocular vision, spinal alignment, and difficulty pumping on a swing. It seems to me that if we are going to specify process dysfunction, then the least we can do is specify the process.

Now, back to the definition. The process disorder "...may manifest itself in an imperfect ability to listen, think, read, write, spell, or do mathematical calculations." Think for a moment about imperfect or perfect ability. Have you ever sat at a presentation and had your mind wander? Have you ever misinterpreted or misheard something? Have you ever made a logical error or drawn an inappropriate conclusion from confusing data? Have you ever misarticulated, spoken with an accent, failed to say what you really meant? Have you ever been writing a paper and had your mind go faster than your pencil and left out words? Can you spell every word in the English langauge? Have you ever made an error in your checkbook? Sounds like the seven warning signals of learning disability doesn't it?

The last part of most definitions excludes students eligible for other kinds of special services. Yet, I think, there is a logical absurdity to the statement that learning disabled children do "not include children who have learning problems which are primarily the result of ...environmental... disadvantage." I think that exclusion fails to take into account several thousand reports on the impact of environment on learning. These and similar empty phrases lead us into a conceptual quagmire. The words and phrases used to define learning disability cannot mean what they literally mean. They've been given new meanings. This phenomenon is generally referred to as "Humpty Dumpty," in honor of the first egghead. Lewis Carroll, in Through the Looking Glass, recorded a conversation between Humpty Dumpty and Alice.

> "When I use a word," Humpty Dumpty said in a rather scornful tone, "it means just what I choose it to mean, neither more nor less." "The question is," said Alice, "whether you can make words mean so many different things." "The question is," said Humpty Dumpty, "which is to be master."

If words in a definition can mean anything we want them to mean, then surely we can use the definition to identify anybody we want to identify. When Kirk and Elkins examined children being served by CSDCs, the elite facilities for LD children, they found that the children actually served under the heading of learning disability would not meet the current definition of learning disability cited in the Federal rules and regulations. One should not be surprised, Humpty Dumpty definitions result in whimsical and capricious procedures; this is the problem that Ysseldyke and Deno face, if they are relying upon such procedures, since they will use LD populations that are already identified. They've got to make some sort of sense out of that conceptual quagmire.

The identification of learning disabled children requires that inadequate definitions be operationalized. We have to do something to turn a definition into practice. I think the nicest thing that we can say about current assessment devices is that they are generally inadequate for diagnostic purposes. But, I think that the time is long since past for us to be so polite. We cannot continue to tolerate such inadequate devices or the continued creation of more inadequate devices.

To conclude, there are some bright spots when considering the practicality of the proposed research. The Deno team offers a potential for immediate help in instructional management and development of IEP's. If their work is successful, I see the likelihood of rapid infusion into current practices by behaviorists. They tend to read, and they tend to use their data. The Ysseldyke team can provide useful descriptions of what is really going on in diagnosis and classification. I think they have the long-term potential to provide information which could be used to reform current diagnostic and classification practices, not only for children with learning disabilities but also for children with mild nonsensory handicaps. I think the Deno team also has the long-term potential, if their learning data prove to be valid and efficient, to provide an educational basis for predicting pupil progress. From my own interests and my own perspectives, I'm satisfied with what the Institute is going to do. At the same time, I would have been overjoyed if these researchers had designed one or two additional studies that demonstrated first, that pupil performance during instruction was more predictive of subsequent diagnosis of learning disability than were psychometric process measures. And second, that teachers using pupil performance data as a basis for instructional decision making were not biased by naturally-occurring stimulus properties for the children they were teaching. I think that if we could have those findings, we could come out of the swamp.

References

Deno, S., & Mirkin, P. _Data based program modification: A manual._ Reston, Virginia: Council for Exceptional Children, 1977.

Highlights of the Roundtable Discussion Concluding the Conference

The discussion began with comments on the blending of the two research projects presented by Ysseldyke and Deno. It was noted that the two projects do seem to have different methodological orientations. It was suggested that an important approach to blending the research would be to get each project to ask the other group's questions while using their own methodology. For example, Deno's team might ask: what kinds of biases occur in decision making when using time-sampling data? Such a blending of questions would provide a type of cross-validation. Also, the two teams might brainstorm the pitfalls of each other's work and possibly how each would conduct the studies proposed by the other team. Methods, to some extent, determine how you work at a problem and even the problems you see. Talking with each other would promote the sharing of insight. The suggestion is not that the two teams do each other's studies, but rather, that each team take into account the important aspects of overlap. For example, Deno's team should protect itself, whenever possible, against biased assessment. And Ysseldyke's team should bring the behavioral approach to bear when looking at the operation of multidisciplinary decision-making teams.

It was further noted that it is probably a very healthy research situation when researchers in an institute do not share a common methodology. The projects in the Institute do not exist in isolation; the Institute setting provides the opportunity to bring individuals together who can challenge one another's ideas. The concern was expressed, however, that it is equally important to synthesize the results of the two approaches, so that the various points of view would be brought together at some point in time.

A question was then asked about how the two research projects in the Institute fit in which reading education. It was suggested that multiple methodologies be used to approach the question of intervention in reading. When one looks at existing programs that are thought to be special, a common factor is a high teacher-pupil ratio. Generally, descriptions of the content or teaching procedure are so gross as to be rather useless. Thus, to help learning disabled children, we must discover the major parameters of the educational programs that help them. A very simple experimental study was recommended, one which would involve essentially planned variation studies and an ethnographic strategy to determine what is going on in programs where children are successful. One could use a naturalistic observation strategy to find out what happens in those classrooms where children make phenomenal headway in reading. If you find that a particular intervention really makes a difference, a difference that can be seen by a regular classroom teacher, then you are in a position to develop an assessment/diagnosis system to locate those types of children so that they can be given the services that you found to be successful for them. This might involve looking for aptitude-treatment interactions (ATI), even though past research using that approach has been very grim.

In commenting on this suggestion, it was noted that ATI research has been grim not only in the past. It is particularly distressing when talking about pupil outcomes. Bruce Joyce at Stanford systematically had teachers use specific instructional approaches and randomly moved children through the various approaches. He was unable to identify any one instructional approach that was effective, any one teacher that was effective across approaches, and any interactions. Such findings suggest that judging outcomes of the teaching process is tremendously complex and perhaps idiosyncratic. Furthermore, even if one were to identify pupil aptitudes that interacted with treatments to produce specific instructional outcomes, everything would wash out if the teacher had a bad day.

Agreement on these points was expressed, but it was also noted that the research to date has focused on a rather narrow set of variables. It was suggested that, in a sense, sending a child to school reflects an ATI decision. Children whom you think have an aptitude to learn are sent to schools; others (for example, hospitalized children so impaired that their chances of learning in a classroom are very low) are not sent to schools. That is a decision that implies an aptitude-treatment interaction. A finer interaction is implied by our decision to send children to school at age six rather than age two, with the assumption that there is more aptitude for learning at six than at two. Even finer decisions, such as placing trainable youngsters in one set of curriculum goals and other children in another set, involve an ATI paradigm. The assumption of interactions in these cases is probably warranted, but data have not been collected at this level. ATI research has been conducted on finer variables, such as the interaction of a divergent-thinker teacher working with children who are divergent thinkers. Maybe the interactions do not exist at that level. But certainly, they do exist at some level or we wouldn't have education.

Discussion then turned to the potential applicability of the formative evaluation approach to virtually any set of goals or targets. The relevant question then becomes: what are the important targets in reading, cognitive development, personal-social development, language, and so on? When these are identified, they can be empirically validated to determine their value within a formative evaluation system. Such a system requires simple definitions of target behaviors; if they are complex, the intensive analytic instructional system cannot be applied because there is a lack of consistent agreement.

The importance of such questions was stressed. It was noted that in reading, for example, comprehension should be a target. Some measures might deal with reading passages and some might deal with processing at the sentence level. The credibility of behavior analysis, outside of the behavior analysis community, will be vastly increased by working on problems that people are concerned with, such as comprehension. In line with this, it was suggested that specifying multiple methods and multiple outcomes becomes critical. How far away are children from their target task? At what point do you get some kind of effect? At what point does your influence end?

The question was posed of why people should expect that teaching a lower level skill, such as word recognition, should affect a higher level skill such as comprehension. Comprehension should not be expected to increase when an intervention program is remediating a lower level process. In reply, it was suggested that a simple-minded approach resulted in such expectations. When a reading comprehension failure is observed, people suggest that the comprehension failure is not surprising because the child does not recognize any of the words. So, the simple approach is to remediate the word recognition problems to increase comprehension. The correct approach probably is to teach comprehension, but that is difficult because we do not know what is involved. Some people have suggested that automaticity is necessary to comprehension, and this seems to be true up to a certain level. But, the area is still quite unclear.

Comments were also made on the possibility of measuring changes in curriculum areas, outside of what is taught specifically in the classroom. It was suggested that it may be mathematically impossible to do that because you would only expect 10 to 20 points of change during a year; that represents only one tenth of a point change each day. Such a change could not be measured reliably every day. Perhaps the only alternative is to track changes in what is being taught specifically in the program. In addition, there seems to be the potential for a higher rate of change the simpler the task being measured. Thus, we find relatively high rates of change for word recognition, but not for comprehension, however comprehension is defined.

Discussion then turned to some cautions that should be observed in the study surveying assessment instruments to determine those that predict success in a systematic instructional program. First, it might be best to crossvalidate the measures discovered for predicting one-year success, by looking at their predictive value after two years and also by seeing whether they predict success with a second group of children. Second, the study should be conducted for more than one systematic instructional program.

It was noted, however, that the results will provide valuable information even if it is possible to demonstrate in only one situation that certain measures do predict success. The most important contribution of the study will be in its interrelationships with the other studies being conducted at the Institute. The results must be looked at in context and as a part of an integrated research endeavor.

Highlights of the Open Discussion Concluding the Conference

Discussion began with comments and questions regarding the extent to which each of the five federally funded Institutes for Research on Learning Disabilities is conducting research on practical problems. Participant observers who are employed in school systems indicated that practical problems must be addressed and solved *now*; school personnel are "under the gun" to develop programs and services today, programs and services that are beneficial to learning disabled children.

While considerable pressure is generated for instant solution of complex enduring problems, it was agreed that practitioners should not rely on the results of inadequate and inconclusive research in formulating services and programs. Equally stressed was the fact that research should be directed toward solution of practical problems and issues, should not be isolated, and should be such that knowledge and results obtained would have direct application in classrooms and school environments.

The Institute was commended for the initial step it has taken in the direction of classroom relevancy by holding the Roundtable Conference with educators in attendance as Roundtable members and as participant-observers. The Institute was encouraged to continue such efforts, perhaps by allowing educators to be involved in the Institute as visiting observers. An important spin-off of this approach might be that these educators could generate related research in their own school systems. It was also suggested that the Institute might work with educators to determine or identify possible appropriate procedures, perhaps based on values of educators and researchers, before all data are in and analyzed. This suggestion was made as a possible approach to the clear need for immediate educational policies, especially as they relate to the placement team decision-making procedure.

A distinction was proposed between applied research and practical research. Although applied research may carry with it the possibility of being practically applicable in school settings, current practical research issues involve such things as implementing Public Law 94-142 and conforming to existing regulations on the evaluation of children with specific learning disabilities. These practical issues are urgent and challenging. The paradigm of conducting research on problems perceived by the schools, rather than those generated by researchers, was supported. It was suggested that the questions being investigated by the Institute may not be those of teachers, parents, and children in the educational system. A relevant research issue, for example, might involve developing and evaluating various assessment models, including the model embodied in the federal regulations, as well as other models that the researchers feel might be of greater efficacy. It was suggested that a "consumer approach" is needed; research should take into account the consumer acceptability of procedures that are to be recommended on the basis of research data. Preferences, rather than data, often determine what procedures will be implemented in the field.

In response to these comments, it was suggested that the problem may not be so much in the research questions asked as in the failure of researchers to be in contact with people in the schools. This does not mean that researchers should necessarily formulate their research questions solely by talking to educators. An example of why this is the case comes from the area of behaviorism. Behavioral research had virtually nothing to do with the practical world for nearly 30 or 40 years, yet in terms of practical applicability today, it is probably the shining glory of psychological research. But, in order to become practically applicable, the researchers had to talk to educators.

It was then noted that there are literally thousands of researchable issues in the field of learning disabilities, most of them tremendously complex. Decisions have to be made about a specific number of questions to ask; such decisions are influenced by parents, advocate agencies, researchers, and a variety of other sources. At this point, people have to accept the fact that only a limited set of issues can be researched at any one time. Everyone is clearly saying that the research should be practical. Issues related to assessment and decision making are critical in education today.

The discussion concluded with comments on the need to keep communication lines open in all directions--to and from educators in the field and to and from other researchers. The comment was made that the Institute was formed to conduct scientific research. Such research must be publicly observable with replicable outcomes. In order to be educationally valuable, the first consideration must be that the research is a scientific enterprise.

PUBLICATIONS

Institute for Research on Learning Disabilities
University of Minnesota

The Institute is not funded for the distribution of its publications. Publications may be obtained for $2 per document, a fee designed to cover printing and postage costs. Only checks and money orders payable to the University of Minnesota can be accepted. All orders must be prepaid.

Requests should be directed to: Editor
Institute for Research on Learning Disabilities
223 Fraser Hall
106 Pleasant St. S.E.
University of Minnesota
Minneapolis, Minnesota 55455

Ysseldyke, J. E. *Assessing the learning disabled youngster: The state of the art* (Research Report No. 1). November, 1977.

*Ysseldyke, J. E. *Research and development plan: University of Minnesota Institute for Research on Learning Disabilities* (Monograph No. 1). June, 1978.

*Ysseldyke, J. E., & Thurlow, M. L. (Eds.). *Synthesis of the knowledge base: Identification and assessment of learning disabled children* (Monograph No. 2). June, 1978.

*Ysseldyke, J. E. *Organizational plan for Minnesota's Institute for Research on Learning Disabilities* (Monograph No. 3). June, 1978.

*Ysseldyke, J. E., & Thurlow, M. L. *Specific investigations to be completed during years two and three* (Monograph No. 4). June, 1978.

*Ysseldyke, J. E., & Thurlow, M. L. *Training opportunities at Minnesota's Institute for Research on Learning Disabilities* (Monograph No. 5). June, 1978.

*Ysseldyke, J. E. *Client base* (Monograph No. 6). June, 1978.

Ysseldyke, J. E., & Regan, R. R. *Nondiscriminatory assessment and decision making* (Monograph No. 7). February, 1979.

Thurlow, M., & Ysseldyke, J. *Assessment in the child service demonstration centers for learning disabled children* (Research Report No. 2). March, 1979.

---

*As part of its continuation proposal, the Institute was required to prepare these monographs. Because they are part of the proposal, they are not available for general distribution.

Foster, G., Algozzine, B., & Ysseldyke, J. Susceptibility to stereotypic bias (Research Report No. 3). March, 1979.

Algozzine, B. An analysis of the disturbingness and acceptability of behaviors as a function of diagnostic label (Research Report No. 4). March, 1979.

Algozzine, B., & McGraw, K. Diagnostic testing in mathematics: An extension of the PIAT? (Research Report No. 5). March, 1979.

Deno, S. L. A direct observation approach to measuring classroom behavior: Procedures and application (Research Report No. 6). April, 1979.

Ysseldyke, J. E., & Mirkin, P. K. Proceedings of the Minnesota roundtable conference on assessment of learning disabled children (Monograph No. 8). April, 1979.

Somwaru, J. P. A new approach to the assessment of learning disabilities (Monograph No. 9). April, 1979.